

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
4 January 2001 (04.01.2001)

PCT

(10) International Publication Number  
WO 01/00858 A1

- (51) International Patent Classification<sup>7</sup>: C12N 15/82, (72) Inventors: JIANG, Jiming; 5734 Schumann Drive, Madison, WI 53711 (US). DONG, Fenggan; 17 Laramie Court, Madison, WI 53706 (US).  
5/10, C12Q 1/68, A01H 5/00
- (21) International Application Number: PCT/US00/17535
- (22) International Filing Date: 26 June 2000 (26.06.2000)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 09/343,775 30 June 1999 (30.06.1999) US
- (71) Applicant: WISCONSIN ALUMNI RESEARCH FOUNDATION [US/US]; P.O. Box 7365, Madison, WI 53707-7365 (US).
- (74) Agent: MUELLER, Lisa, V.; Rockey, Milnamow & Katz, Ltd., Suite 4700, Two Prudential Plaza, 180 North Stetson Avenue, Chicago, IL 60601 (US).
- (81) Designated States (*national*): AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian

[Continued on next page]

(54) Title: DNA SEQUENCES SPECIFIC TO RICE CENTROMERES

10	20	30	40	50
GATCTTTGGA TTGGAACAG TTAAAGAACA ATATGTCAT GATGATGATT				
60	70	80	90	100
TTAAAGATG GTTTTTCAT TATAAGGATG GGAAGGCATG GAATAAATTT				
110	120	130	140	150
CTTGTAATG ATGTTTTGT GTTAGAGCT AATAAGCTAT GCATTCCAGC				
160	170	180	190	200
TAGCTCTGTT CGTTTGTGT TGTACAGCA AGCACAATGA GGTGGTTTCA				
210	220	230	240	250
TGGGACATTT TGGGCAAG AGACGGAGG ACATCTGGC TGCTCATTTT				
260	270	280	290	300
TPTTGCCAA AGATGAGGAG AGATGTGGAG AGATTATTG CTCGTCGAC				
310	320	330	340	350
GACATGTCAA AAGGCCAAGT CACGCTTAA TCCACAGAT TTGAAGCCAT				
360	370	380	390	400
ATTGGGTGA GGGAGATGAG CTGAGTCGG GGACGACTCA AATGCAAGAA				
410	420	430	440	450
GGGAGGATG ATGAGSACAT CAGCACCATC TATACATCCA CACCTACACC				
460	470	480	490	500
CACACCATCG CCAACACCAC TTGGCCCTCT TACTCTGCC AGTGCCCGTC				
510	520	530	540	550
AACTGAACCA TCAAGTAAGT TTATTCTTAA ACTTTGTCC ATCATATTTA				
560	570	580	590	600
GACAATGGAG ACACGTGCAC TCTTGTTTG CTTAGGAATG ATGGAGAGGA				
610	620	630	640	650
CCAGAAGCAT AGGGATTGG TGTAGGCTGG ATTGGACAG CAAGACAGCA				
660	670	680	690	700
CCAACTTACA ACAACCGCCA TGACTTCATA CAGAGTCCAT TTAAAGCATG				
710	720	730	740	750
CAAGCACTTG ATGGAAACT GGTCAAATAT ATTTTATAGT GGATC				

[Continued on next page]



WO 01/00858 A1



patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**Published:**

— With international search report.

— Before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

---

(57) Abstract: The present invention relates to nucleic acids which encode a functional centromere from *Oryza sativa*. The nucleic acids of the present invention can be also used to create a plant artificial chromosome.

## DNA SEQUENCES SPECIFIC TO RICE CENTROMERES

This invention was made with United States government support awarded by the  
5 following agencies: USDA HATCH 3935. The United States has certain rights in this invention.

Field of the Invention

The present invention relates generally to molecular biology. In particular, the present  
invention relates to nucleic acid sequences which encode a centromere from rice.

10

Background of the Invention

Among the most distinguishing and characteristic landmarks of chromosomes of higher  
eukaryotes is the location of the centromere. The centromeric region is the site for mitotic and  
meiotic spindle fiber attachment and is responsible for sister chromatid association. Jiang, J., et  
15 al., *Proc. Natl. Acad. Sci. USA*, 95: 8135-8140 (1998). Therefore, centromeres play a central role  
in the process of chromosomal segregation and transmission in cell divisions. *Id.* The molecular  
organization of centromeres has been studied extensively in yeast, *Drosophila melanogaster*, and  
humans.

20 Centromeric regions usually consist of heterochromatin and are thought to be highly  
methylated. Miller, J.T., et al., *Theor. Appl. Genet.*, 96:832-839 (1998). In addition, centromeres  
show varying amounts of nontranscribed repetitive sequences, which are referred to as satellite  
DNAs. Haaf, T., et al., *Cell*, 70:681-695 (1992). The predominant class of centromeric DNA, is  
the alpha-satellite DNA, which is found in diverged form in all centromeres. *Id.* To the extent  
25 that it is known, alpha-satellite arrays appear to be uninterrupted by other (nonsatellite) DNA  
sequences.

In centromeres, naturally occurring satellite arrays range in size from several hundred kbs  
to several megabases in length. Recent studies, however, suggest that as little as 140 kb of alpha  
30 satellite DNA may be sufficient to confer centromere function in human cells. Harrington, John.

J., et al., *Nature Genetics*, 15:345-355 (1997). Unfortunately, satellite DNA of this size has proven difficult to clone and propagate stably in microorganisms using conventional cloning vectors. *Id.* In large part, the difficulty in propagating satellite DNA stems from the tendency of tandemly repetitive DNA to recombine into smaller arrays and this effect increases with the size of the repetitive array. *Id.*

As briefly mentioned hereinbefore, functional centromeric sequences have been isolated and purified from *S. cerevisiae* (see Clark et al., *Nature*, 287:504-509 (1980) and Stinchcomb et al., *J. Molec. Biol.* 158:157-179 (1982)). Episomes carrying the yeast centromeric sequences display proper segregation into daughter yeast cells during mitosis and meiosis, in contrast to autonomous replication sequences plasmids lacking a centromere.

The best characterized centromeric DNAs originated from the budding yeast *S. cerevisiae* (See Clarke and Carbon, *Ann. Rev. Genet.* 19:29-56 (1985)). The DNA region required for centromere function in *S. cerevisiae* is approximately 120 base pair (hereinafter "bp") long and is composed of three conserved domains: CDEI, an 8 bp element (A/G)TCAC(A/G)TG), CDEII, an extremely (about 90%) AT-rich region of approximately 80 bp, and CDEIII, a 25 bp element (TGTTT(A/T)TGNTTTCGAAANNNA). The molecular structure of centromeric DNAs from the fission yeast *Schizosaccharomyces pombe* have also been characterized. Several classes of *S. pombe* moderately repeated DNA elements have been identified which are found only in the centromere regions. These centromere-specific repetitive elements have been designated dg (3.8 kb), dh (4 kb), and yn by Yanagida and co-workers (Nakaseko et al., *Embo. J.* 5:1011-1021 (1986); Nakaseko et al., *Nuc. Acid Res.* 15:4705-4715 (1987)), and K (6.4 kb), L (6 kb), and B (1 kb) by Carbon and his colleagues (Clarke et al., *PNAS* 83:8253-8257 (1986); Fishel et al., *Mol. Cell Biol.* 8:754-763 (1988)). The dg element has an AT-rich region and a 600 bp domain containing numerous small direct repeat motifs. Similarly, the dh element has an overall AT content approaching 70% and contains many short direct repeats. No nucleotide similarities to the *S. cerevisiae* CDEs have been found in the *S. pombe* elements.

Attempts to demonstrate that the *S. pombe* centromere-specific repetitive elements can function individually as centromeres have been unsuccessful. However, large restriction fragments (65 to 150 kb) carrying the entire fission yeast centromere regions of chromosome 1 or 3 function as centromeres when introduced into acentric episomes (Hahnenberger et al., *PNAS USA* 86:577-581 (1989)). These results indicate that either fission yeast centromeres are large composite structures that cannot be subdivided, or the functional fission yeast centromere element has not yet been identified.

In contrast to the detailed studies done in *S. cerevisiae* and *S. pombe*, in most eukaryotes, only limited information is available regarding the organization of the centromeres. For example, limited information is known about plant centromeres. Peacock et al., *Proc. Natl. Acad. Sci. USA*, 78:4490-4494 (1981) report the first isolation of a repetitive DNA element from maize knobs. This repetitive DNA element acts as neocentromeres in certain genetic backgrounds. A repetitive DNA element has also been cloned from the centromeres of the supernumerary B chromosomes of maize (see Alfenito, M.R., et al., *Genetics* 135:589-597 (1993) and Kaszas, E., et al., *EMBO J.*, 15:5246-5255 (1996)). Part of this B-specific DNA element shows strong homology to the maize sequences. A 180-bp tandem repeat (pAL1 family) is the major component of the centromeric region of *Arabidopsis thaliana* chromosomes. The genomic organization of this repeat family shares similarities to the alpha satellite DNA at the human centromeres (see Martinez-Sapater, J., et al., *Mol. Gen. Genet.*, 204:417-423 (1986); Simoens, C. R., *Nucleic Acids Res.*, 16:6753-6766 (1988); Maluszynska, J., et al., *Plant J.*, 1:159-166 (1991); Round, E.K., et al., *Genome Res.*, 7:1045-1053 (1997)).

As discussed above, very few putative functional centromeres have been cloned from plants. The cloning of a putative functional centromere from a plant is a necessary first step in the production of artificial chromosomes suitable for use in plants. Artificial chromosomes are man-made linear or circular DNA molecules constructed from essential cis-acting DNA sequence elements that are responsible for the proper replication and partitioning of natural chromosomes (see, Murray et al., *Nature*, 305:189-193 (1983)). The essential elements of an artificial

chromosome are: Autonomous Replication Sequences (ARS) (have properties of replication origins, which are the sites for initiation of DNA replication), (2) centromeres (site of kinetochore assembly and responsible for proper distribution of replicated chromosomes at mitosis and meiosis), and (3) telomeres (specialized structures at the ends of linear chromosomes that function to stabilize the ends and facilitate the complete replication of the extreme termini of the DNA molecule). The use of artificial chromosomes as an alternative to commonly used method of introducing new genetic information into cells is steadily increasing.

#### Summary of the Invention

The present invention relates to isolated and purified nucleic acids having the nucleotide sequences shown in: SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6 and SEQ ID NO:7.

The present invention also relates to a recombinant DNA construct which contains a rice centromere. The centromere contains a number of highly repetitive regions of DNA that have the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6 and SEQ ID NO:7 or combinations thereof. The recombinant DNA construct may also contain a yeast autonomous replication sequence, an autonomous replication sequence from a higher eukaryotic organism, a yeast telomere sequence or a telomere sequence from a higher eukaryotic organism and a selectable marker gene.

The present invention also relates to a plasmid containing the hereinbefore described DNA construct. This plasmid may contain an origin of replication and a selectable marker which functions in bacteria (such as *E. coli*) or in yeast (such as *S. cerevisiae*).

The present invention relates to a plant artificial chromosome vector. The plant artificial chromosome vector of the present invention contains an autonomous replication sequence, two telomere sequences, a centromere sequence having the nucleotide sequence of SEQ ID NO:1,

SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6 and SEQ ID NO:7 or a combination thereof, and at least one selectable marker sequence. The autonomous replication sequence may be from yeast or from a higher eukaryotic organism and the telomere sequence may be from yeast or from a higher eukaryotic organism, such as, but not limited to, *Arabidopsis thaliana*.

The present invention also relates to a plant cell transformed with the plant artificial chromosome vector hereinbefore described and to transgenic plants containing said plant cell. The plant cell and plant may be from *Oryza sativa*.

Finally, the present invention relates to a method of identifying centromeric DNA in a higher eukaryotic organism. The method involves hybridizing an isolated nucleic acid selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6 and SEQ ID NO:7 and combinations thereof with a sample of DNA from a higher eukaryotic organism and then identifying and isolating the centromeric DNA from said sample.

#### Brief Description of the Drawings

FIG. 1 shows the nucleotide sequence of pSau3A9.

FIG. 2A - FIG. 2N shows the FISH analysis of rice centromeric DNA elements. The probes were biotinylated and hybridized *in situ* to rice chromosomes or DNA fibers. The probes were detected by fluorescein isothiocyanate-conjugated antibodies (green color) and the chromosomes were stained with propidium iodide (red color). Probes pRCS1 hybridized exclusively to the centromeric regions of the chromosomes from rice (FIG. 2A), rye (FIG. 2B), barley (FIG. 2C), sorghum (FIG. 2D), and maize (FIG. 2E). FISH signals also were detected in the centromeric regions of the acrocentric B chromosomes (see arrows) from rye (FIG. 2B) and maize (2E). Similarly, rice centromeric DNA families RCH2 (FIG. 2F), RCH1 (FIG. 2G), RCH3

(FIG. 2H), RCE1 (FIG. 2I), RCE2 (FIG. 2J), and RCS2 (FIG. 2K) all were located in the centromere of every rice chromosome. Two pairs of chromosomes with the strongest signals are indicated by arrows and the third pair with the weakest signals by arrowheads (FIG. 2K). The same metaphase cell (FIG. 2K) was washed under medium (FIG. 2L) and high stringencies (FIG. 2M), and most signals were still discernible (FIG. 2N). The marked array between two arrows is 51  $\mu$ m long and represents approximately 151-kb DNA. All bars are 10  $\mu$ m.

FIG. 3 shows a Southern blot of the genomic organization of the RCS1 family. Rice genomic DNA was digested with *Sau3AI* (lane 1), *DpnII* (lane 2), *HaeIII* (lane 3), *MspI* (lane 4), *HpaII* (lane 5), *SalI* (lane 6), *BamHI* (lane 7), *DraI* (lane 8), *EcoRI* (lane 9), and *HindIII* (lane 10) and probed with pRCS1.

FIG 4 shows the nucleotide sequence of pRCS2. The 639-bp insert of pRCS2 contains four copies of a tandemly arranged repeat. The four members (A-D) range from 155 to 165 bp and share 84-91% sequence identity with one another. F represents the consensus sequence of the four members.

FIG. 5 shows a Southern blot of the genomic organization of the RCS2 family. Rice genomic DNA was digested with *DpnII* (lane 1), *Sau3AI* (lane 2), *MspI* (lane 3), *HpaII* (lane 4), *SalI* (lane 5), and *HaeIII* (lane 6), and probed with pRCS2.

FIG. 6A - FIG. 6B show a Southern blot of the conservation of the RCH1 and RCE1 families in *Gramineae* species. Genomic DNA from sorghum (lane 1), maize (lane 2), sugar cane (lane 3), *Ag. intermedium* (lane 4), barley (lane 5), oats (lane 6), rye (lane 7), wheat (lane 8), *Ae. Squarrosa* (lane 9), rice (lane 10), bamboo (lane 11), *Pharus sp.* (lane 12), *J. effusus* (lane 13), *C. alternifolius* (lane 14) and *A. thaliana* (lane 15) was digested with *HindIII* and probed with pRCH1 (FIG. 6A) and pRCE1 (FIG. 6B).

FIG. 7 shows the nucleotide sequence of RCS1.



FIG. 8 shows the nucleotide sequence of RCS2.

FIG. 9 shows the nucleotide sequence of RCH1.

5 FIG. 10 shows the nucleotide sequence of RCH2.

FIG. 11 shows the nucleotide sequence of RCH3.

FIG. 12 shows the nucleotide sequence of RCE1.

10 FIG. 13 shows the nucleotide sequence of RCE2.

### Detailed Description of the Invention

#### 15 Background

The present invention relates to cloned centromeric DNA from *Oryza sativa* (rice). More specifically, the inventors of the present invention have discovered that the cloned centromeric DNA of the present invention contains seven (7) different repetitive regions of complex DNA. These seven (7) repetitive regions are referred to herein as follows: RCS1, RCS2, RCH1, RCH2, 20 RCH3, RCE1 and RCE2.

The present invention relates to isolated and purified nucleic acids for each of the seven (7) different repetitive regions of centromeric DNA from *Oryza sativa*. The nucleic acids of the present invention encode a functional centromere from *Oryza sativa*.

25 The present invention further relates to the use of the nucleic acids of the present invention as primers and probes to identify centromeric DNA from other plants and animals. In addition, the nucleic acid sequences disclosed herein can be used to create a plant artificial chromosome vector.

### Definitions

Units, prefixes, and symbols can be denoted in the SI accepted form. Numeric ranges are inclusive of the numbers defining the range. Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation, respectively. The headings provided herein are not  
5 limitations of the various aspects or embodiments of the invention which can be had by reference to the specification as a whole. Accordingly, the terms defined immediately below are more fully defined by reference to the specification as a whole.

As used herein, the term "plant" includes reference to whole plants, plant organs (e.g.,  
10 leaves, stems, roots, etc.), seeds and plant cells and progeny thereof. The class of plants which can be used in the methods of the present invention are generally as broad as the class of higher plants amenable to transformation techniques, including both monocotyledonous and dicotyledonous plants.

As used herein, the term "transformation" or "transfection" means the acquisition in cells  
15 of new DNA sequences through incorporation of added DNA. This is the process by which naked DNA, DNA coated with protein, or whole artificial chromosomes are introduced into a cell, resulting in a heritable change.

As used herein, the term "host" means any organism that is the recipient of a replicable  
20 plasmid or vector comprising a plant artificial chromosome. Preferably, host strains used for cloning are free of any restriction enzyme activity that might degrade the foreign DNA used. Preferred examples of host cells for cloning which are useful in the present invention are bacteria, such as *Escherichia coli*, *Bacillus subtilis*, *Pseudomonas*, *Streptomyces*, *Salmonella*,  
25 and yeast cells such as *S. cerevisiae*. Host cells which can be targeted for expression of a plant artificial chromosome may be plant cells of any source, such as, but not limited to, *Arabidopsis*, maize, rice, sugarcane, sorghum, barley, soybeans, tobacco, wheat, tomato, potato or citrus.

As used herein, the term "linker" means a DNA molecule, generally up to 50 or 60

nucleotides long and synthesized chemically, or cloned from other vectors.

As used herein, the term "plasmid" or "vector" (such as a cloning vector or expression vector)" refers to a closed covalently circular extrachromosomal DNA or linear DNA which is able to autonomously replicate in a host cell and which is normally nonessential to the survival of the cell. A wide variety of plasmids and other vectors are well known and commonly used in the art.

As used herein, "heterologous" when used to describe nucleic acids or polypeptides refers to nucleic acids or polypeptides that originate from a foreign species, or, if from the same species, are substantially modified from their original form. For example, a promoter operably linked to a heterologous structural gene is from a species different from that from which the structural gene was derived, or, if from the same species, one or both are substantially modified from their original form.

As used herein, "isolated" includes reference to material which is substantially or essentially free from components which normally accompany or interact with it as found in its naturally occurring environment. The isolated material optionally comprises material not found with the material in its natural environment.

As used herein, "nucleic acid" includes reference to a deoxyribonucleotide or ribonucleotide polymer in either single- or double-stranded form, and unless otherwise limited, encompasses known analogues of natural nucleotides that hybridize to nucleic acids in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence includes the complementary sequence thereof.

As used herein, "operably linked" includes reference to a functional linkage between a promoter and a second sequence, wherein the promoter sequence initiates and mediates transcription of the DNA sequence corresponding to the second sequence. Generally, operably

linked means that the nucleic acid sequences being linked are contiguous and, where necessary to joint two protein coding regions, contiguous and in the same reading frame.

5 As used herein "recombinant" includes reference to a cell, or nucleic acid, or vector, that has been modified by the introduction of a heterologous nucleic acid or the alteration of a native nucleic acid to a form not native to that cell, or that the cell is derived from a cell so modified. For example, recombinant cells express genes that are not found within the native (non-recombinant) form of the cell or express native genes that are otherwise abnormally expressed, under expressed or not expressed at all.

10

As used herein, a "recombinant DNA construct" is a nucleic acid construct, generated recombinantly or synthetically, with a series of specified nucleic acid elements which permit transcription and translation of a particular nucleic acid in a target cell. The DNA construct can be part of a plasmid, vector, virus, or nucleic acid fragment. Typically, the recombinant DNA  
15 construct portion of the construct includes a nucleic acid to be transcribed and translated, and a promoter. In the present invention, the recombinant DNA construct can be a plant artificial chromosome.

20

As used herein, "transgenic plant" includes reference to a plant modified by introduction of a heterologous nucleic acid.

As used herein, "telomere" refers to the end of a chromosome comprising a simple repeat DNA. The function of a telomere is to allow the ends of a linear DNA molecule to be replicated.

25

As used herein, "eukaryote" refers to living organisms whose cells contain nuclei. A eukaryote may be distinguished from a "prokaryote" which is an organism which lacks nuclei. Prokaryotes and eukaryotes differ fundamentally in the way their genetic information is organized, as well as their patterns of RNA and protein synthesis.

As used herein, "lower eukaryote" refers to a eukaryote characterized by a comparatively simple physiology and composition, and unicellularity. Examples of lower eukaryotes include flagellates, ciliates, and yeast.

5 As used herein, "higher eukaryote" refers to a multicellular eukaryote, characterized by its greater complex physiological mechanisms as well as its ability to interact with its environment in a more sophisticated manner. Generally, more complex organisms such as plants and animals are included in this category. Preferred higher eukaryotes to be transformed by the present invention include, for example, monocot and dicot angiosperm species, gymnosperm  
10 species, fern species, plant tissue culture cells of these species, and algal cells. It will of course be understood that prokaryotes and eukaryotes alike may be transformed by the methods of this invention.

As used herein, a "selectable marker" is a gene whose presence results in a clear  
15 phenotype, and most often a growth advantage for cells that contain the marker. This growth advantage may be present under standard conditions, altered conditions such as elevated temperature, or in presence of certain chemicals such as herbicides or antibiotics. Examples of selectable markers include the thymidine kinase gene, the cellular adenine-  
phosphoribosyltransferase gene and the dihydrofolate reductase gene, hygromycin  
20 phosphotransferase genes, the bar gene and the neomycin phosphotransferase genes, among others. Preferred selectable markers in the present invention include genes whose expression confer antibiotic or herbicide resistance to the host cell, sufficient to enable the maintenance of a vector with a host cell, and which facilitate the manipulation of a plasmid into new host cells.

25 As used herein, "nucleotide" refers to one of the monomeric units from which DNA or RNA polymers are constructed, consisting of a purine or pyrimidine base, a pentose, and a phosphoric acid group. The nucleotides of DNA are deoxyadenylic acid, thymidylic acid, deoxyguanylic acid, and deoxycytidylic acid. The corresponding nucleotides of RNA are adenylic acid, uridylic acid, guanylic acid, and cytidylic acid.

## SEQUENCE LISTING

The present application also contains a sequence listing that contains 8 sequences. The sequence listing contains nucleotide sequences. For the nucleotide sequences, the base pairs are represented by the following base codes:

	<u>Symbol</u>	<u>Meaning</u>
	A	A; adenine
	C	C; cytosine
10	G	G; guanine
	T	T; thymine
	U	U; uracil
	M	A or C
	R	A or G
15	W	A or T/U
	S	C or G
	<u>Symbol</u>	<u>Meaning</u>
	Y	C or T/U
20	K	G or T/U
	V	A or C or G; not T/U
	H	A or C or T/U; not G
	D	A or G or T/U; not C
	B	C or G or T/U; not A
25	N	(A or C or G or T/U)

Nucleic Acid Sequences

In one embodiment, the present invention relates to isolated and purified nucleic acids which encode a functional centromere from *Oryza sativa*. As used herein, the term "a functional centromere" refers to the centromere or chromosome site that directs or supports kinetochore formation. The kinetochore is the physical structure that mediates the attachment of the spindle fibers to the chromosome and is therefore responsible for the proper partition of the chromosomes at mitosis and meiosis.

The nucleic acids of the present invention encode seven (7) different repetitive regions of centromeric DNA from *Oryza sativa*. Exemplary nucleic acids for such centromeres have the nucleotide sequences shown in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6 and SEQ ID NO:7 or combinations thereof. SEQ ID NO:1 is also referred to herein as RCS1. SEQ ID NO:2 is also referred to herein as: RCS2. SEQ ID NO:3 is also referred to herein as RCH1. SEQ ID NO:4 is also referred to herein as RCH2. SEQ ID NO:5 is referred to herein as RCH3. SEQ ID NO:6 is referred to herein as RCE1. SEQ ID NO:7 is referred to herein as RCE2.

The present invention also contemplates nucleic acids which hybridize under stringent hybridization conditions to the nucleotide sequences set forth above. Generally, stringent conditions are selected to be about 5°C to about 20°C lower than the thermal melting point (T<sub>m</sub>) for the specific sequence at a defined ionic strength and pH. The T<sub>m</sub> is the temperature (under defined ionic strength and pH 7) at which 50% of the target sequence hybridizes to a perfectly matched probe. Typically, stringent wash conditions are those in which the salt concentration is about 0.22 molar at pH 7 and the temperature is at least about 50°C. However, nucleic acids which do not hybridize to each other under stringent conditions are still substantially identical if it encodes a substantially identical and functional centromere. This may occur, e.g., when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. The present invention also contemplates naturally occurring allelic variations and mutations of the nucleotide sequences set forth above so long as those variations and mutations code, on expression, for a functional centromere.

As is well known in the art, because of the degeneracy of the genetic code, there are numerous other DNA and RNA molecules that can code for the same functional centromere as encoded by SEQ ID NOS: 1-7, or portions thereof. The present invention, therefore, contemplates those other DNA and RNA molecules, which, on expression, encode for a functional centromere encoded by the nucleic acid sequences of SEQ ID NOS: 1-7 or portions thereof. With knowledge of all triplet codons for each particular amino acid residue, it is

possible to describe all such encoding RNA and DNA sequences. DNA and RNA molecules other than those specifically disclosed herein and, which molecules are characterized simply by a change in a codon for a particular amino acid are within the scope of this invention. A table of codons representing particular amino acids is set forth below in Table 1.

TABLE 1

First Position (5' end)	Second Position				Third Position (3' end)
	T/U	C	A	G	
T/U	Phe	Ser	Tyr	Cys	T/U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	Stop	Stop	A
	Leu	Ser	Stop	Stop	G
C	Leu	Pro	His	Arg	T/U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	T/U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	T/U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

The nucleic acid sequences of the present invention can be used in marker-aided selection using techniques which are well-known in the art. Marker-aided selection does not require the complete sequence of the gene or precise knowledge of which sequence confers which specificity. Instead, partial sequences can be used as hybridization probes or as the basis for



oligonucleotide primers to amplify by PCR or other methods to identify nucleic acids specific for functional centromeric DNA in other plants and animals.

### Plant Artificial Chromosome

5 In a second embodiment, the present invention relates to a plant artificial chromosome. More specifically, the nucleic acid sequences of the present invention can be used to construct a plant artificial chromosome vector. A plant artificial chromosome must contain the following essential elements: (1) autonomous replication sequences (hereinafter referred to as "ARS"), (2) a centromere which is functional in plants, and (3) telomeres which are functional in plants.

### Autonomous Replication Sequences

10 ARSs have been isolated from the unicellular fungi *Saccharomyces cerevisiae* (brewer's yeast) and *Schizosaccharomyces pombe* (see Stinchcomb et al., *Nature* 282:39-43 (1979) and Hsiao et al., *J. Proc. Natl. Acad. Sci. USA* 76:3829-3833 (1979)). ARSs behave like replication  
15 origins allowing DNA molecules that contain the ARS to be replicated as an episome after introduction into the cell nuclei of these fungi. Although plasmids containing these sequences replicate, they do not segregate properly.

20 U.S. Patent 5,270,201 (hereinafter the "201 Patent"), hereby incorporated by reference, discloses a method for isolating ARS sequences for use in higher eukaryotic organisms, by the formation of minichromosomes derivative of natural chromosomes. It has been demonstrated in yeast that inverted repeats of telomeric sequences are "resolved" by an unknown mechanism which results in a double-stranded cleavage between inverted repeats. After an inverted telomere repeat is introduced into a chromosome, a resolution reaction will lead to scission of the  
25 chromosome and formation of two chromosomal fragments, each with two telomeres. This process generates a minichromosome small enough to be isolated intact allowing further manipulation by *in vitro* techniques to delimit the sequences required for autonomous replication.

A second approach for an obtaining ARS is also disclosed in the '201 Patent. This approach is referred to as a "shotgun cloning approach". Higher eukaryotic organisms have many replication origins distributed throughout their genomes. For example, the *A. thaliana* genome contains approximately 1000 origins spaced every 70 kb along the chromosome.

5 Therefore, the shotgun cloning approach involves looking for random fragments of genomic DNA throughout the genome of interest which promote extrachromosomal replication.

Autonomous replication sequences for use in the plant artificial chromosome of the present invention can be obtained using methods which are well known in the art. Autonomous  
10 replication sequences from yeast, such as those described above, can be used in the present invention. Moreover, ARS sequences from higher eukaryotic organisms obtained using the methods described in the '201 patent can also be used in plant artificial chromosome of the present invention.

#### 15 Telomeres

Telomeres are believed to be involved in the priming of DNA replication at the chromosome end (see, Blackburn et al., *Ann. Rev. Biochem.* 53:163-194 (1984)). This is because conventional DNA polymerases are template dependent, synthesize DNA in the 5' to 3' direction, and require an oligonucleotide primer to donate a 3' OH group. When this primer is removed,  
20 unreplicated single-stranded gaps arise; most of these gaps can be filled in by priming from 3' OH groups donated by newly replicated strands located at the 5' end of the gap. However, the unreplicated gaps which lie next to the extreme 5' end of the DNA duplex cannot be primed in this manner. Consequently, telomeres must provide an alternative priming mechanism.

25 Telomeres are also responsible for the stability of chromosomal termini. Telomeres act as "caps," suppressing the recombinogenic properties of free, unmodified DNA ends (see Blackburn et al., *Ann. Rev. Biochem.* 53:163-194 (1984)). This reduces the formation of damaged and rearranged chromosomes which arise as a consequence of recombination-mediated chromosome

fusion events.

Telomeres may also contribute to the establishment or maintenance of intranuclear chromatin organization through their association with the nuclear envelope (see, for example, 5 Fussell, C. P., *Genetica* 62:192-201 (1984)).

Telomeric or telomeric-like DNA sequences have been cloned from several lower eukaryotic organisms, principally protozoans and yeast. The ends of the *Tetrahymena* linear DNA plasmid have been shown to function like a telomere on linear plasmids in *S. cerevisiae* 10 (see Szostak, J. W., Cold Spring Harbor Symp. *Quant. Biol.* 47:1187-1194 (1983)). A telomere from the flagellate *Trypanosoma* has been cloned (see, for example, Blackburn et al., *Cell* 36:447-457 (1984)). A yeast telomeric sequence has been identified (see, for example, Shampay et al., *Nature* 310:154-157 (1984)).

15 U.S. Patent 5,270,201 disclose a method for obtaining a telomere from a higher eukaryotic organism, specifically, from *Arabidopsis thaliana*. The telomeric sequences disclosed in the '201 Patent contain a tandem repeat of the sequence 5'-CCCTAAA-3.

20 Any telomeric sequence which produces a telomere which is functional in plants can be inserted into the plant artificial chromosome of the present invention. The telomeric sequence may be from yeast or from a higher eukaryotic organism as described above. Preferably, the plant artificial chromosome of the present invention will contain two (2) telomeric sequences.

#### Construction of a Plant Artificial Chromosome

25 Once the essential elements of a plant artificial chromosome are obtained (the ARS, centromere and telomeres), a plant artificial chromosome vector can be constructed using methods which are well-known in the art (see, for example, Maniatis, T., et al., *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor, 1982)).

In addition to the essential elements described above, preferably positive and negative selectable plant markers (for example, antibiotic or herbicide resistance genes), and a cloning site for insertion of foreign DNA are preferably included. In order to propagate vectors in *E. coli*, it is necessary to convert the linear molecule into a circle by the addition of a stuffer fragment  
5 between the telomeres. In addition to the stuffer fragment, the artificial plant chromosome may also contain a origin of replication that can function in plants.

Artificial plant chromosomes which replicate in yeast also may be constructed to take advantage of the large insert capacity and stability of repetitive DNA inserts afforded by this  
10 system (Burke et al., *Science*, 236:806-812 (1987)). In this case, yeast ARS and centromere sequences are added to the artificial chromosome. The artificial chromosome is maintained in yeast as a circular molecule using a stuffer fragment to separate the telomeres.

Nucleic acids for the essential components of the plant artificial chromosome obtained  
15 from any source whatsoever, may be purified and inserted into the plant artificial chromosome at any appropriate restriction endonuclease cleavage site. The nucleic acids usually will contain various regulatory signals (for example, promoters, termination segments, enhancers, etc., which are well known in the art) that allow for the expression of proteins encoded by the nucleic acids. Alternatively, regulatory signals residing in the artificial chromosome may be utilized.

20 The techniques and procedures required to accomplish insertion are well-known in the art (see Maniatis *et al.*, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., (1982)). Typically, this is accomplished by incubating a circular plasmid or a linear DNA fragment in the presence of a restriction endonuclease such that the  
25 restriction endonuclease cleaves the DNA molecule. Endonucleases preferentially break the internal phosphodiester bonds of polynucleotide chains. They may be relatively unspecific, cutting polynucleotide bonds regardless of the surrounding nucleotide sequence. However, the endonucleases which cleave only a specific nucleotide sequence are called restriction enzymes. Restriction endonucleases generally internally cleave DNA molecules at specific recognition

sites, making breaks within "recognition" sequences that in many, but not all, cases exhibit two-fold symmetry around a given point. Such enzymes typically create double-stranded breaks.

Many of these enzymes make a staggered cleavage, yielding DNA fragments with protruding single-stranded 5' or 3' termini. Such ends are said to be "sticky" or "cohesive" because they will hydrogen bond to complementary 3' or 5' ends. As a result, the end of any DNA fragment produced by an enzyme, such as *EcoRI*, can anneal with any other fragment produced by that enzyme. This properly allows splicing of foreign genes into plasmids, for example. Some restriction endonucleases that may be particularly useful with the current invention include *HindIII*, *PstI*, *EcoRI*, and *BamHI*.

Some endonucleases create fragments that have blunt ends, that is, that lack any protruding single strands. An alternative way to create blunt ends is to use a restriction enzyme that leaves overhangs, but to fill in the overhangs with a polymerase, such as klenow, thereby resulting in blunt ends. When DNA has been cleaved with restriction enzymes that cut across both strands at the same position, blunt end ligation can be used to join the fragments directly together. The advantage of this technique is that any pair of ends may be joined together, irrespective of sequence.

Those nucleases that preferentially break off terminal nucleotides are referred to as exonucleases. For example, small deletions can be produced in any DNA molecule by treatment with an exonuclease which starts from each 3' end of the DNA and chews away single strands in a 3' to 5' direction, creating a population of DNA molecules with single-stranded fragments at each end, some containing terminal nucleotides. Similarly, exonucleases that digest DNA from the 5' end or enzymes that remove nucleotides from both strands have often been used. Some exonucleases which may be particularly useful in the present invention include *Bal31*, *SI*, and *ExoIII*. These nucleolytic reactions can be controlled by varying the time of incubation, the temperature, and the enzyme concentration needed to make deletions. Phosphatases and kinases also may be used to control which fragments have ends which can be joined. Examples of useful

phosphatases include shrimp alkaline phosphatase and calf intestinal alkaline phosphatase. An example of a useful kinase is T4 polynucleotide kinase.

5       Once the source DNA sequences and vector sequences have been cleaved and modified to generate appropriate ends they are incubated together with enzymes capable of mediating the ligation of the two DNA molecules. Particularly useful enzymes for this purpose include T4 ligase, *E. coli* ligase, or other similar enzymes. The action of these enzymes results in the sealing of the linear DNA to produce a larger DNA molecule containing the desired fragment (see, for example, U.S. Pat. Nos. 4,237,224; 4,264,731; 4,273,875; 4,322,499 and 4,336,336, which are  
10       specifically incorporated herein by reference).

      It is to be understood that the termini of the linearized plasmid and the termini of the DNA fragment being inserted must be complementary or blunt in order for the ligation reaction to be successful. Suitable complementarity can be achieved by choosing appropriate restriction  
15       endonucleases (*i.e.*, if the fragment is produced by the same restriction endonuclease or one that generates the same overhang as that used to linearize the plasmid, then the termini of both molecules will be complementary). As discussed previously, in a preferred embodiment, at least two classes of the vectors used in the present invention are adapted to receive the foreign oligonucleotide fragments in only one orientation. After joining the DNA segment to the vector,  
20       the resulting hybrid DNA can then be selected from among the large population of clones or libraries.

      A method useful for the molecular cloning of DNA sequences includes *in vitro* joining of DNA segments, fragmented from a source high molecular weight genomic DNA, to vector DNA  
25       molecules capable of independent replication. The cloning vector may include plasmid DNA (see Cohen *et al.*, *Proc. Natl. Acad. Sci. USA*, 70:3240 (1973)), phage DNA (see Thomas *et al.*, *Proc. Natl. Acad. Sci. USA*, 71:4579 (1974)), SV40 DNA (see Nussbaum *et al.*, *Proc. Natl. Acad. Sci. USA*, 73:1068 (1976)), yeast DNA, *E. coli* DNA and most significantly, plant DNA.

A variety of processes are known which may be utilized to effect transformation; *i.e.*, the inserting of a heterologous DNA sequences into a host cell, whereby the host becomes capable of efficient expression of the inserted sequences.

5        Transformed Host Cells and Transgenic Plants

Methods and compositions for transforming a bacterium, a yeast cell, a plant cell, or an entire plant with one or more plant artificial chromosome vectors are further aspects of the present invention.

10        Means for transforming bacteria and yeast cells are well known in the art. Typically, means of transformation are similar to those well known means used to transform other bacteria or yeast such as *E. coli* or *Saccharomyces cerevisiae*. Methods for DNA transformation of plant cells include *Agrobacterium*-mediated plant transformation, protoplast transformation, gene transfer into pollen, injection into reproductive organs, injection into immature embryos and  
15        particle bombardment. There are various advantages and disadvantages associated with each of these methods.

Methods for transforming plant cells include any method by which DNA can be introduced into a cell, such as by *Agrobacterium* infection, direct delivery of DNA such as, by  
20        PEG-mediated transformation of protoplasts, by desiccation/inhibition-mediated DNA uptake, by electroporation, by agitation with silicon carbide fibers, by acceleration of DNA coated particles, etc.

Many methods for delivering genes into cells are known and well described. These  
25        methods include: (1) chemical methods (Graham et al., *Virology*, 54(2):536-539 (1973); Zatloukal et al. *Ann. N.Y. Acad. Sci.*, 660:136-153 (1992)); (2) physical methods such as microinjection (Capecchi, *Cell* 22(2):479-488 (1980)), electroporation (Wong et al., *Biochim. Biophys. Res. Commun.*, 107(2):584-587 (1982); Fromm et al., *Proc. Natl. Acad. Sci. USA*, 82(17):5824-5828 (1985); U.S. Patent 5,384,253) and microprojectile bombardment (*i.e.* the

gene gun) (Johnston et al., *Methods Cell. Biol.*, 43(A):353-365 (1994); Fynan et al., *Proc. Natl. Acad. Sci. USA*, 90(24):11478-11482 (1993); (3) viral vectors (Clapp, *Clin. Perionatol.*, 20(1):155-168 (1993); Lu et al., *J. Exp. Med.*, 178(6):2089-2096 (1993)); Eglitis et al. *Biotechniques*, 6(7):608-614 (1988); Eglitis et al. *Avd. Exp. Med. Biol.*, 241:19-27 (1988); and  
5 (4) receptor-mediated mechanisms (Curiel et al., *Proc. Natl. Acad. Sci. USA*, 88(19):8850-8854 (1991); Curiel et al., *Hum. Gen. Ther.*, 3(2):147-154 (1992); Wagner et al., *Proc. Natl. Acad. Sci. USA* 89(13):6099-6103 (1992)).

*Agrobacterium*-mediated transformation is a widely applicable system for introducing  
10 genes into plant cells because the DNA can be introduced into whole plant tissues. The use of *Agrobacterium*-mediated plant integrating vectors to introduce DNA into plant cells is well known in the art. Using conventional transformation vectors, chromosomal integration is required for stable inheritance of the foreign DNA. However, the artificial plant chromosome vector described herein may be used for transformation with or without integration, as the  
15 centromere function required for stable inheritance is encoded within the plant artificial chromosome. In particular embodiments, transformation events in which the plant artificial chromosome is not chromosomally integrated may be preferred, in that problems with site-specific variations in expression and insertional mutagenesis may be avoided.

20 The integration of the Ti-DNA is a relatively precise process resulting in few rearrangements. The region of DNA to be transferred is defined by the border sequences, and intervening DNA is usually inserted into the plant genome as described (Spielmann et al., *Mol. Gen. Genet.*, 205:34 (1986); Jorgensen et al., *Mol. Gen. Genet.*, (1987)). Modern *Agrobacterium* transformation vectors are capable of replication in *E. coli* as well as *Agrobacterium*, allowing  
25 for convenient manipulations as described (Klee et al., *Bio/Technology*, 3:637-642(1985)). Moreover, recent technological advances in vectors for *Agrobacterium*-mediated gene transfer have improved the arrangement of genes and restriction sites in the vectors to facilitate construction of vectors capable of expressing various polypeptide coding genes. The vectors described (Rogers et al., *Meth. In Enzymol.*, 153:253-277 (1987)), have convenient multi-linker



regions flanked by a promoter and a polyadenylation site for direct expression of inserted polypeptide coding genes and are suitable for present purposes. In addition, *Agrobacterium* containing both armed and disarmed Ti genes can be used for the transformations. In those plant strains where *Agrobacterium*-mediated transformation is efficient, it is the method of choice because of the facile and defined nature of the gene transfer.

*Agrobacterium*-mediated transformation of leaf disks and other tissues such as cotyledons and hypocotyls appears to be limited to plants that *Agrobacterium* naturally infects.

*Agrobacterium*-mediated transformation is most efficient in dicotyledonous plants. Few monocots appear to be natural hosts for *Agrobacterium*, although transgenic plants have been produced in asparagus and more significantly in maize using *Agrobacterium* vectors as described (Bytebier *et al.*, *Proc. Natl. Acad. Sci. USA*, 84:5345 (1987)); U.S. Patent No. 5,591,616, specifically incorporated herein by reference). Therefore, commercially important cereal grains such as rice, corn, and wheat must usually be transformed using alternative methods. However, as mentioned above, the transformation of asparagus using *Agrobacterium* also can be achieved (see, for example, Bytebier *et al.*, *Proc. Natl. Acad. Sci. USA*, 84:5345 (1987)).

#### Other Transformation Methods

Transformation of plant protoplasts can be achieved using methods based on calcium phosphate precipitation, polyethylene glycol treatment, electroporation, and combinations of these treatments (see, for example, Potrykus *et al.*, *Mol. Gen. Genet.* 199:183 (1985); Lorz *et al.*, *Mol. Gen. Genet.*, 199:178 (1985); Fromm *et al.*, *Nature*, 312:791-793 (1986); Uchimiya *et al.*, *Mol. Gen. Genet.*, 204:204 (1986); Callis *et al.*, *Genes and Development*, 1:1183 (1987); Marcotte *et al.*, *Nature*, 335:454 (1988)).

Application of these systems to different plant strains for the purpose of making transgenic plants depends upon the ability to regenerate that particular plant strain from protoplasts. Illustrative methods for the regeneration of cereals from protoplasts are described

(Fujimura *et al.*, *Plant Tissue Culture Letters*, 2:74 (1985); Toriyama *et al.*, *Theor. Appl. Genet.*, 73:16 (1986); Yamada *et al.*, *Plant Cell Rep.*, 4:85 (1986); Abdullah *et al.*, *Biotechnology*, 4:1087 (1986)).

5 To transform plant strains that cannot be successfully regenerated from protoplasts, other ways to introduce DNA into intact cells or tissues can be utilized. For example, regeneration of cereals from immature embryos or explants can be effected as described (Vasil, *Biotechnology* 6:397 (1988)). In addition, "particle gun" or high-velocity microprojectile technology can be utilized (Vasil *et al.*, *Biotechnology*, 10:667-674 (1992)).

10 Using that latter technology, DNA is carried through the cell wall and into the cytoplasm on the surface of small metal particles as described (Klein *et al.*, *Nature*, 327:70-73 (1987); Klein *et al.*, *Proc. Natl. Acad. Sci. USA*, 85:8502-8505 (1988); McCabe *et al.*, *Biotechnology*, 6:923 (1988)). The metal particles penetrate through several layers of cells and thus allow the  
15 transformation of cells within tissue explants.

By way of example, and not of limitation, Examples of the present invention will now be given.

## 20 **EXAMPLE 1: CLONING OF CENTROMERIC DNA FROM *ORZYA SATIVA***

### **1. Materials and Methods.**

a. A rice BAC library was constructed from an indica rice (*Oryza sativa* ssp. Indica) line IR-BB21 and consisted of 11,000 clones (see Wang, G.-L., *et al.*, *Plant J.* 7: 525-533 (1995),  
25 herein incorporated by reference). The cereal centromeric DNA element pSau3A9 (Jiang, J., *et al.*, *Proc. Natl. Acad. Sci. USA* 93:14210-14213 (1996), herein incorporated by reference) was used to isolate the rice centromere-specific BAC clones. The DNA sequence of pSau3A9 is shown in Figure 1 and SEQ ID NO:8. Rice lines used in this example include a javanica rice (*O. sativa* ssp. Javanica) line DV85, a japonica rice (*O. sativa* ssp. Japonica ) line Norin 28, and

indica rice line IR72, and four other *Oryza* species (*O. glaberrima*, *O. rufipogon*, *O. officinalis*, and *O. alta*). Gramineae species used in conservation studies include two species from the Bambusoideae subfamily [bamboo (*Bambusa vulgaris*), *Pharus* sp.], three species from the Panicoideae subfamily [sorghum, maize (*Zea mays*), and sugar cane (*Saccharum officinarum*)], six species from the Pooideae subfamily [*Agropyron intermedium*, barley (*Hordeum vulgare*), oats (*Avena sativa*), rye (*Secale cereale*), wheat (*Triticum aestivum*), and *Aegilops squarrosa*]. Three non-Gramineae species, *Juncus effusus*, *Cyperus alternifolius*, and *A. thaliana*, and rye and maize lines containing B chromosomes also were included.

**b. BAC Library Screening.** BAC filter preparations and BAC library screening were conducted as described in Wang, G.-L., et al., *Plant J.* 7:525-533 (1995); Hoheisel, J. D., et al., *Cell* 73:109-120 (1993), herein incorporated by reference. BAC clones were isolated by using pSau3A9 as a probe, and their cytological locations were confirmed by fluorescence *in situ* hybridization (hereinafter referred to as "FISH").

**c. Subcloning and Sequencing.** DNA fragments recovered from agarose gels were subcloned into pUC18 plasmids as described in Jiang, J., et al., *Proc. Natl. Acad. Sci. USA* 93: 14210-14213 (1996), herein incorporated by reference. Cycle sequencing reactions were performed by using Applied Biosystems AmpliTaq DNA polymerase, FS Dye Terminator Ready Reactions kit, and a Perkin-Elmer Thermocycler (model 2400). Reaction products were analyzed on an Applied Biosystems DNA sequencer (model 373).

**d. Southern Blot Hybridization.** Plant genomic DNA was isolated as described in Gill, K. S., et al., *Genome* 34:362-374 (1991), herein incorporated by reference. BAC DNA was prepared by using an alkaline lysis method described in Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed., pp. 1.25-1.26, herein incorporated by reference, and purified by CsCl ultracentrifugation. Gel transfers, prehybridizations, hybridizations, and posthybridization

washing were all as previously described Jiang, J., et al., *Proc. Natl. Acad. Sci. USA* 93: 14210-14213 (1996).

c. **Slot Blot Hybridization.** Copy number of each subclone in rice genome was determined by slot blot hybridization as described Zhao, X., et al., *Theor. Appl. Genet.* 78:201-209 (1989), herein incorporated by reference. Band intensities were measured on the autoradiographs by IPLab Spectrum v3.1 software.

f. **FISH.** Detailed protocols for FISH and Fiber-FISH are described in Jiang, J., et al., *Proc. Natl. Acad. Sci. USA* 93: 14210-14213 (1996) and Fransz, P. F., et al., *Plant J.* 9:421-430 (1996), herein incorporated by reference. The formamide in the hybridization mixture was 50% and 30% in regular and low stringency hybridizations, respectively. Washing was conducted at either low [ $2 \times$  saline sodium citrate (SSC) at 42°C for 15 minutes], medium (50% formamide at 45°C for 15 minutes) or high stringency (70% formamide at 50°C for 15 minutes).

2. A rice BAC library constructed from indica rice (*Oryza sativa* ssp. Indica) and described by Wang, G.-L., et al., *Plant J.* 7: 525-533 (1995), was screened by using pSau3A9 as a probe. Twenty-two clones showed unambiguous positive hybridizations. Ten of the 22 clones were analyzed cytologically by FISH. Eight clones hybridized to the centromeric or/and paracentromeric regions of all rice chromosomes. Clone 17p22 showed bright and sharp signals specific to the centromeric regions. At a low hybridization stringency, this clone also hybridized exclusively to the centromeric regions of chromosomes from sorghum, maize, wheat, barley, oats and rye.

DNA from clone 17p22 was digested with restriction enzymes, *Bam*HI, *Dra*I, *Eco*RI, *Hae*III, *Hind*III, *Msp*I, *Pst*I, *Sau*3AI and *Sal*I and blotted onto nylon membrane. Small DNA fragments ranging from 0.5 to 3 kb were subcloned, and their distinctiveness was confirmed by Southern hybridization using blots containing 17p22 DNA digested with the above described nine restriction enzymes. Seven different DNA families, including two *Sau*3AI fragments

(subclones pRCS1 and pRCS2), three *Hind*III fragments (subclones pRCH1, pRCH2 and pRCH3), and two *Eco*RI fragment (subclones pRCE1 and pRCE2), were identified (see below in Table 2). These seven families hybridized to all of the fragments generated by the nine enzymes. FISH and Southern hybridization analysis indicated that all seven elements are repetitive in the rice genome (see below).

**Table 2**  
**Summary of the Seven Rice Centromeric Repetitive DNA Families**

Family	SEQ ID NO:	Size, bp	GC content, %	Organization pattern	Copy number*	Conservation
RCS1	1	1478	40	Dispersed	130	Gramineae
RCS2	2	639	41	Tandem	6,200#	<i>Oryza</i>
RCH1	3	827	45	Dispersed	53	Gramineae
RCH2	4	1,201	46	Dispersed	99	Gramineae
RCH3	5	1,341	48	Dispersed	67	Gramineae
RCE1	6	701	39	Dispersed	287	Bambusoideae
RCE2	7	2,863	41	Dispersed	305	Gramineae

\*Based on the haploid genome of rice as 424 Mb (24).

#The copy number of the 168-bp monomer in the rice genome.

Clone pRCS1 contains a 877-bp *Sau*3AI fragment that hybridizes to the pSau3A9 sequence. Sequencing analysis revealed that the 259 bp at the 3' end of pRCS1 had 80% sequence identity to the central part (bases 338-602) of the pSau3A9 sequence (see Jiang, J., et al., *Proc. Natl. Acad. Sci. USA* 93: 14210-14213 (1996), FIG. 1 and SEQ ID NO:8). The first 95 bp in pRCS1 had 76% sequence identity to a Ty3/*gypsy* class of retrotransposon sequence reported in maize (GenBank Accession No. AF030633). Nucleotides 171-228 of pRCS1 had 70% sequence identity to a Ty3/*gypsy* class of retrotransposon sequence reported in *Lilium*

*henryi* (X13886). It was also discovered that the pSau3A9 sequence in sorghum has similar sequence identities to the Ty3/*gypsy* class of retrotransposons. These results indicated that both pSau3A9 and pRCS1 probably were derived from retrotransposon-related DNA sequences.

5        The RCS1 sequence was located in the centromeric regions of all 24 rice chromosomes by FISH (see FIG. 24). The sizes and intensities of the FISH signals were uniform on different chromosomes, suggesting that all rice chromosomes contain a similar number of copies of this element. Slot blot analysis suggested that there are about 130 copies of RCS1 present in the haploid genome of japonica rice DV85 (see Table 2).

10        Rice genomic DNA was digested with several restriction enzymes and probed with the 259-bp fragment conserved between rice and sorghum. One or few major bands and several minor bands were detected in most of the lanes (see FIG. 3). Fiber-FISH using pRCS1 as a probe did not generate clustered signals. These results suggested that the RCS1 sequence is dispersed  
15        in the centromeric regions of rice chromosomes.

20        FISH analysis revealed that pRCS1 also hybridized exclusively to centromeric regions of chromosomes from other *Gramineae* species (see FIG. 2B-E). The FISH results on rye (see FIG. 2B) and barley (see FIG. 2C) chromosomes showed that hybridization was exclusive to the primary constrictions. FISH signals also were detected in the centromeres of the supernumerary B chromosomes from both rye and maize (see FIG. 2B and E). Positive Southern hybridization signals were detected in all other *Gramineae* species analyzed, including bamboo, *Pharus* sp., oats, wheat, sugar cane, *Ae. squarrosa*, and *Ag. intermedium*. However, homologous sequences could not be detected by Southern hybridization analysis in dicot species and any monocot  
25        species outside of *Gramineae*, indicating that the RCS1 family is sufficiently conserved only in the grass family *Gramineae*.

      Clone pRCS2 contains a 639-bp *Sau3AI* fragment consisting of four copies of a tandemly arranged repeat with a consensus sequence of 168bp (see FIG. 4). The four copies were 84-91%

identical with one another. The third copy of the repeat contains a 6-bp insertion (TTGGCC) at base 147. A search of the GenBank database found a highly significant match to a repetitive DNA element isolated from *O. Sativa* (GenBank Accession No. U63977).

5           Southern hybridization analysis of rice genomic DNA using probe pRCS2 revealed ladder patterns using several restriction enzymes, including *DpnII*, *Sau3AI*, *MspI*, *HpaII*, and *HaeIII*, indicating that the RCS2 family is tandemly arranged in the rice genome (see FIG. 5). Several enzymes produced digestion profiles comprised of monomer and multiples (dimer, trimer, tetramer, etc.) of the 168-bp basic repeat.

10           Probe pRCS2 hybridized only to the centromeric regions on all rice chromosomes (see FIG. 2K). Significant variation in the size and intensity of the FISH signals was detected in different centromeres. Two pairs of chromosomes had strong signals, and a third pair had very faint signals (see FIG. 2K). All of the signals became weaker as the posthybridization washing stringency was increased (see FIG. 2L). However, even after washing in 70% formamide at 15           50°C for 15 minutes, most signals were still discernible (see FIG. 2M), suggesting that the signal disparity reflects difference in copy numbers rather than sequence divergence of the RCS2 family in different rice centromeres. Though the longest chromosome (chromosome 1) had the strongest signals, it was not possible to relate the copy numbers to the chromosome sizes. It was 20           evident that the weakest signals were not on the smallest chromosomes (see FIG. 2K).

          Three subspecies of *O. sativa* (AA genome), together with *O. glaberrima* (AA), *O. rufipogon* (AA), *O. alta* (CCDD), and *O. officinalis* (CC) were included for FISH analysis. FISH signals were detected in the centromeric regions from all of the chromosomes of these species. 25           Southern hybridization analysis revealed that the RCS2 family is present only in the species within genus *Oryza*. Homologous sequences could not be detected even at a low stringency in any plant species outside of genus *Oryza*.

RCS2 is the most abundant element isolated from BAC 17p22 and has about 1,550 copies, corresponding to 6,200 monomers, in the haploid genome of DV85 (see Table 2 above). BAC 17p22 contains about 46 copies of this element, corresponding to approximately 39% of the BAC insert. Fiber-FISH analysis demonstrated that the RCS2 family is organized into various sizes of uninterrupted arrays in the rice genome. The longest observed block with small interspersed gaps ( $<2 \mu\text{m}$ ) was  $51 \mu\text{m}$  (see FIG. 2N). Based on a  $2.96\text{-kb}/\mu\text{m}$  resolution of the Fiber-FISH technique, this block represents approximately 151 kb of uninterrupted RCS2 sequences. The longest observed single Fiber-FISH signal with interspersed gaps larger than  $2 \mu\text{m}$  was  $188 \mu\text{m}$  representing approximately 556 kb of centromeric DNA sequences.

The other five centromeric DNA elements isolated from rice BAC clone 17p22 were analyzed by FISH, and all of them hybridized exclusively to the centromeric regions of all rice chromosomes (see FIG. 2F-J). One or two pairs of rice metaphase chromosomes showed weak hybridization when pRCH2, pRCH3, and pRCE2 were used as probed. No relationship can be confirmed between signal intensities and the sizes of the chromosomes.

The sequence information for these families is listed in Table 2, above. Searching in the GenBank database did not uncover any significant matches to these sequences except for pRCH2. Bases 39-102 and 204-232 in pRCH2 had sequence identities to the centromeric CCS1 sequence isolated from *B. sylvaticum* (see Aragon-Alcaide, L., et al., *Chromosoma* 105, 261-268 (1996); Abbo, S., et al., *Chromosome Res.* 3:5-15 (1995)). Interestingly, about 120 bp (bases 8-130) of this element had 80% sequence identity to the spacer sequence that separates the rice 5S rRNA genes. The possibility that this element associates with the 5S rDNA locus was excluded because the FISH signals from pRCH2 was located proximal to those from the 5S rDNA locus.

In Southern hybridization analysis, all five elements produced one or few major bands and several minor bands under several restriction enzymes, similar to the RCS1 family (see FIG. 3), suggesting that they all are dispersed in the rice centromeric regions. The copy numbers of these elements ranged from 53 to 305 copies per haploid rice genome (see Table 2, above).



All five elements were hybridized to various plant species by Southern hybridization. The RCE1 family was present only in the species from the Bambusoideae subfamily, including rice, bamboo, and *Pharus* sp. (See FIG. 6B), whereas RCH1, RCH2, RCH3, and RCE2 all were conserved across the Gramineae species (see FIG. 6A for RCH1). Species from subfamily  
5 Panicoideae and Bambusoideae generally had stronger hybridization signals than those from subfamily Pooideae (see FIG. 6A).

The cytosine nucleotides, especially those in dinucleotide sequence 5'CpG3', are the most common sites for methylation in plant genomes. Methylation occurs at lower frequencies when  
10 the C and G are separated by 1-2 A/T nucleotides (see Gruenbaum, T., et al. *Nature (London)* 292: 860-862 (1981)). Enzymes *MspI* and *HpaII* are isoschizomers that recognize the 5'CCGG3' sequence. Neither enzyme can cut when the 5'C is methylated, and only *MspI* can cleave when the internal cytosine is methylated. Though both enzymes produced similar digestion profiles of rice genomic DNA, *MspI* generated much smaller-sized hybridization bands from all of the rice  
15 centromeric DNA probes than *HpaII* did (see FIG. 3 for RCS1 and FIG. 5 for RCS2). For the RCS2 element, monomers of the 168-bp basic repeat could be found in *MspI* lane, and most of the hybridization was in the fragments smaller than 2 kb, whereas the majority of hybridization in the *HpaII* lane was larger than 2 kb (see FIG. 5). For the other centromeric elements, DNA fragments smaller than 5 kb were not detected in *HpaII* lanes (see FIG. 3 for RCS1). These  
20 results suggest that the cytosine of the CpG dinucleotides are heavily methylated in the rice centromeric DNA sequences. Restriction enzyme *SalI* recognizes 5'GTCGAC3' and is sensitive to the methylation of CpG dinucleotides. Small fragments (<10 kb) that hybridized to the centromeric elements were not detected in the *SalI* lanes (see FIGS. 3 and 5).

What is Claimed Is:

1. An isolated and purified nucleic acid comprising a nucleotide sequence of SEQ ID NO:7.
- 5 2. A recombinant DNA construct comprising a centromere, wherein said centromere comprises a number of highly repetitive regions of DNA having a nucleotide sequence of SEQ ID NO:7.
- 10 3. The recombinant DNA construct of claim 2, further comprising a yeast autonomous replication sequence.
4. The recombinant DNA construct of claim 2, further comprising an autonomous replication sequence from a higher eukaryotic organism.
- 15 5. The recombinant DNA construct of claim 2, further comprising a yeast telomere.
6. The recombinant DNA construct of claim 2, further comprising a telomere from a higher eukaryotic organism.
- 20 7. The recombinant DNA construct of claim 2, further comprising a selectable marker gene.
8. A plasmid comprising the DNA construct of claim 2.
- 25 9. The plasmid of claim 8, wherein said plasmid further comprises an origin of replication and a selectable marker that functions in bacteria.
10. The plasmid of claim 9, wherein said bacteria is *E. coli*.
- 30 11. The plasmid of claim 8, wherein said plasmid further comprises an origin of replication and a selectable marker that functions in yeast.

12. The plasmid of claim 8, wherein said yeast is *S. cerevisiae*.
13. A plant artificial chromosome vector comprising an autonomous replication sequence, two telomere sequences, a centromere sequence having the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6 and SEQ ID NO:7 or combinations thereof, and at least one selectable marker sequence.
14. The plant artificial chromosome vector of claim 13, wherein the autonomous replication sequence is from yeast.
15. The plant artificial chromosome vector of claim 13, wherein the autonomous replication sequence is from a higher eukaryotic organism.
16. The plant artificial chromosome vector of claim 13, wherein the telomere sequences are from a higher eukaryotic organism.
17. The plant artificial chromosome of claim 16, wherein the telomere sequences are from *Arabidopsis thaliana*.
18. The plant artificial chromosome vector of claim 13, wherein the telomere sequences are from yeast.
19. A plant cell transformed with the plant artificial chromosome vector of claim 13.
20. The transformed plant cell of claim 19, wherein the plant cell is from *Oryza sativa*.
21. A transgenic plant comprising the transformed plant cell of claim 19.
22. A method of identifying centromeric DNA in a higher eukaryotic organism, the method comprising the steps of:

hybridizing an isolated nucleic acid selected from the group consisting of SEQ  
ID NO:1, SEQ ID NO:2, SEQ ID:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID  
NO:6, SEQ ID NO:7 and combinations thereof, with a sample of DNA from a  
higher eukaryotic organism; and identifying and isolating centromeric DNA  
from said sample.

5

10

15

20

25

30

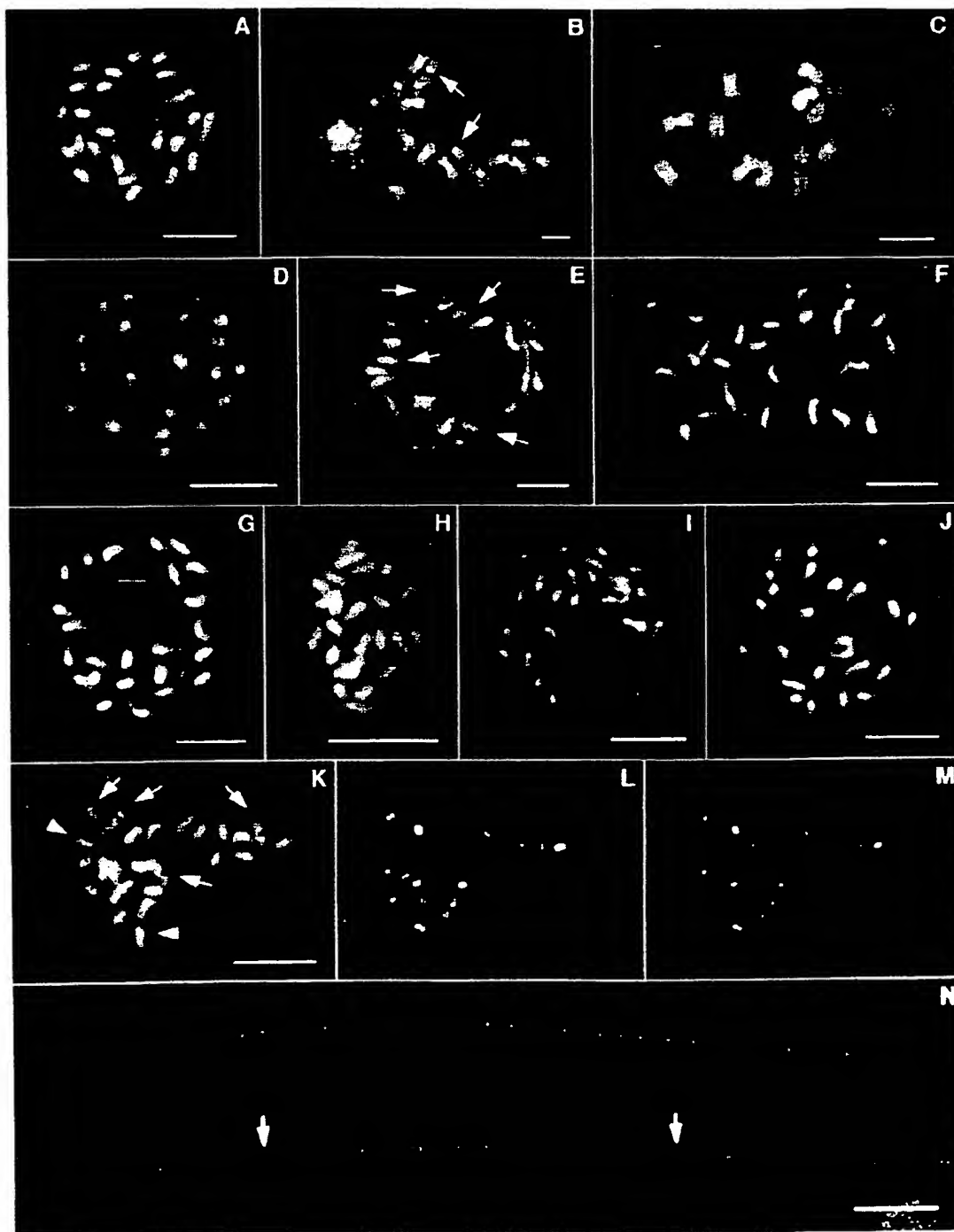
FIG. 1

1/13

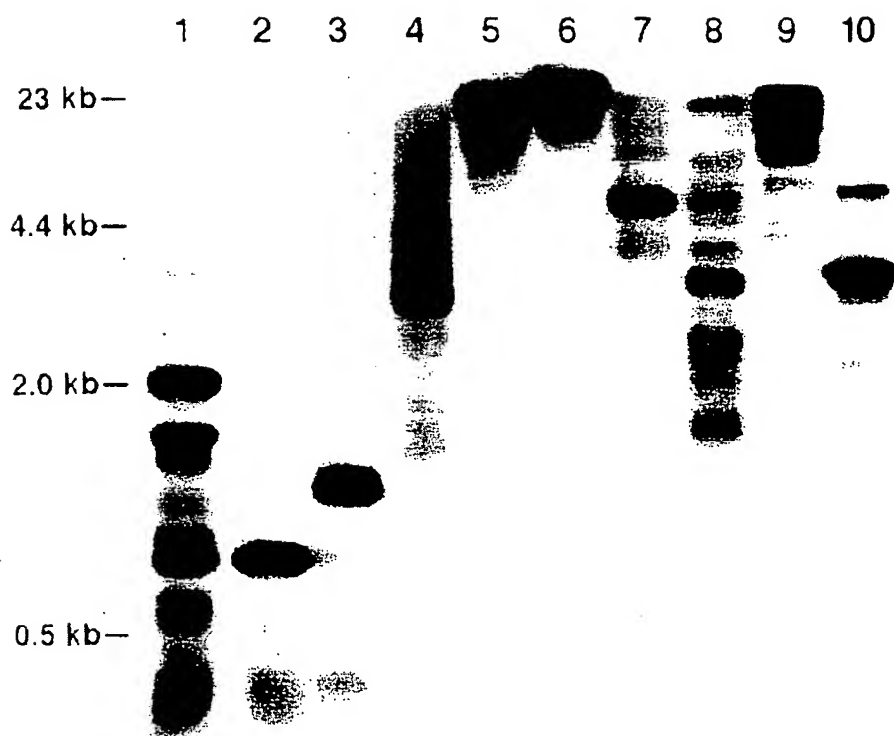
10	20	30	40	50
GATCTTTGGA TTGGAAACAG TTAAAGAACA ATATGTGCAT GATGATGATT				
60	70	80	90	100
TTAAAGATGT GTTTTTGCAT TGTAAGGATG GGAAGGCATG GAATAAATTT				
110	120	130	140	150
GTTGTAAATG ATGGTTTTGT GTTTAGAGCT AATAAGCTAT GCATTCCAGC				
160	170	180	190	200
TAGCTCTGTT CGTTTGTTGT TGCTACAGGA AGCACATGGA GGTGGTTTGA				
210	220	230	240	250
TGGGACATTT TGGGGCAAAG AAGACGGAGG ACATACTGGC TGGTCATTTT				
260	270	280	290	300
TTTTGGCCAA AGATGAGGAG AGATGTGGAG AGATTTATTG CTCGCTGCAC				
310	320	330	340	350
GACATGTCAA AAGGCCAAGT CACGCTTAA TCCACACGAT TTGAAGCCAT				
360	370	380	390	400
ATTTGGGTGA GGGAGATGAG CTTGAGTCGG GGACGACTCA AATGCAAGAA				
410	420	430	440	450
GGGGAGGATG ATGAGGACAT CAGCACCATC TATACATCCA CACCTACACC				
460	470	480	490	500
CACACCATCG CCAACACCAC TTGGCCCTCT TACTCGTGCC AGTGCCCGTC				
510	520	530	540	550
AACTGAACCA TCAAGTAAGT TTATTCTTAA ACTCTTGTCC ATCATATTTA				
560	570	580	590	600
GACAATGGAG ACACGTGCAC TCTTGTTTTG CTTAGGAATG ATGGAGAGGA				
610	620	630	640	650
CCAGAAGCAT AGGGGATTGG TGTAGGCTGG ATTTGGACAG CAAGACAGCA				
660	670	680	690	700
CCAACTTACA ACAACCGCCA TGACTTCATA CAGAGTCCAT TTTAAGCATG				
710	720	730	740	750
CAAGCACTTG ATGGAAAAT CGTCAAGTAT ATTTTGTAGAT GGATC				

2/13

Fig. 2



3/13

**Fig. 3**

4/13

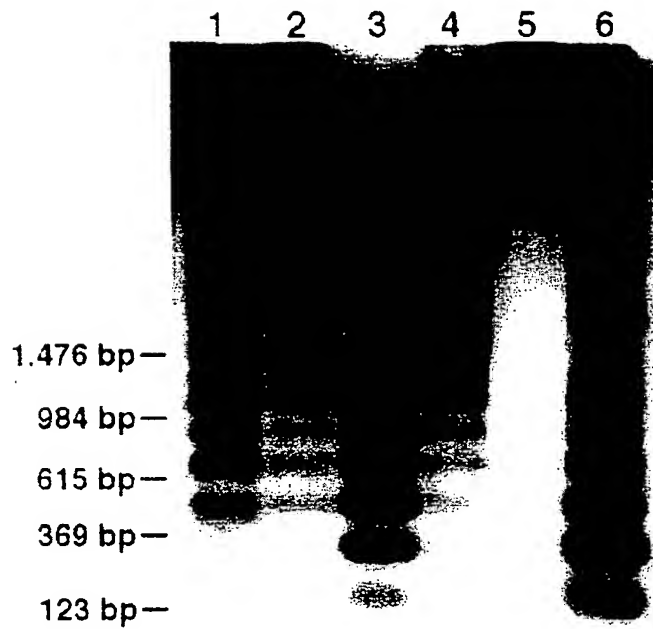
## FIG. 4

	1				50
A	GATC	TT-TTCTACT	-GGAATCAAA	ATGTTCAAAA	AATGCCAAAA CATGATTTTTT
B		TT-TTCTACT	-GGAATCAAA	ATGTTCAAAA	AGAGCCAAAA CATGATTTTTT
C		TTATTCTACT	-GGAATCAAA	ATGTTCAAAA	AGAGCCAAAA CATGATTTTTT
D		TTA-T-TAAT	CGGAATCAAA	ATGTTCAAAA	GGCACCAAAA CATGATTTTTT
F		TTatTcTAcT	-GGAATcAAA	AaGTTCAAAA	agagCCAAAA CATGATTTTTT
	51				100
A		GGACATATTG	GAGTGTATTG	GGTGCATTCA	TGGCAAAAAC TCACTCCGTG
B		GGACATATTG	GAGTGTATTG	GGTGC GTTCA	TGGCAAAA-C TCACTTCGTG
C		GGACATATTG	GAGTGTATTG	GGTGC GTTTCG	TGGCAAAA-C TCACTTCGTG
D		TGACATATTG	GAGTGTATTG	GGTGC GTTTCG	TGGCAAAAAC TCACTTCGTG
F		gGACATATTG	GAGTGTATTG	GGTGCgTTCg	TGGcAAAAaC TCACTtCGTG
	101				150
A		ATTCGCGCGG	CGAACTTTTG	TCAATTAATG	CCAATAT-TG GG-ACA----
B		ATTCGCGCGG	CGAACTTTTG	TCATTTAATG	CCAATATGTG CATACA----
C		ATTCGCGCGG	CGAACTTTTG	TCAATTAAT	CCAATATGTG CATATTTTGG
D		ATTCGCGCGG	CGAACTTTTG	TCATTTAATG	CCAATAT-TG GC-ACA---G
F		ATTCGCGCGG	CGAACTTTTG	TCA*TTAATg	CCAATATgTG *atAca---g
	151		168		
A		--CGAG-G-G	T-GCGATG	(155 bp)	
B		--CGAGAGAG	T-GCGATG	(158 bp)	
C		CCCAAA-GTG	TTGCGATG	(165 bp)	
D		--CGA-CGGG	T-GCGATC	(157 bp)	
F		--CgAg*G*G	T GCGATg		

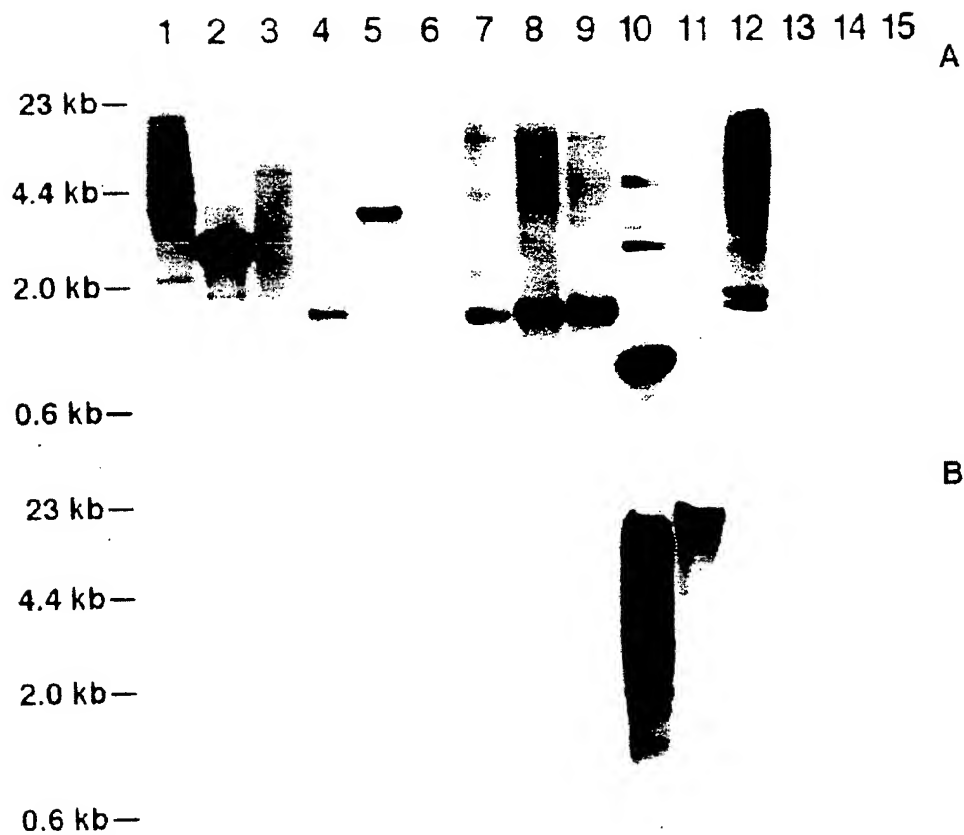


5/13

**Fig. 5**



6/13

**Fig. 6**

7/13

FIG. 7

1	tcataaatga	tgggtttgtt	ttcagagcta	acaagctatg	cattccagct	agctccgttc
61	gcttgttgtt	gttgcaggaa	gcgcattgag	gcgatttgat	gggacatttt	ggtgccaaaga
121	agacacatga	catccttgct	agtcatttct	t-ttggccaca	gatgcgaaga	gatgttggca
181	ggttcgttgc	tcgctacgct	acatgtcaaa	aggctaagtc	acgcttacat	ccacatgggt
241	tgtatatgcc	tcttctctgt	cctactgttc	cttgggaaga	tatttcaatg	gatttttgtgt
301	taggattgcc	taggaccaag	agggggcgctg	atagcatttt	tgtggttgctg	gacgattttt
361	ctaaaaatggc	acattttcata	ccatgtcata	aaactgatga	tgccttctcat	atcgctgatc
421	tgttcttttcg	agaaatttgt	cgcttgcatg	gtgtgccaaa	cacaatttgt	tctgattcgtg
481	acacaaaatt	tcttagccat	ttttggagaa	ctttgtgggc	taaattgggg	actaaaacttt
541	tgttttctac	tacttgtcat	ccccaaactg	atggacaaac	tgaagtgggtg	aatagaacct
601	tgtctactat	gcttagggct	gtttgaaaga	aaalatcaag	atgtgggaag	aatgcttgcc
661	tcataattgaa	tttgcttata	atcgtttcctt	gcatttclact	acaaaaaatg	tgcccatctc
721	agatttgtga	tggtttgttg	cctcgtgctc	caattgattt	gatgccttta	ccatcttctg
781	agaaactgaa	ttttgatgctg	aagcaacgtg	ctgagttgat	gttaaaactg	catgagacaa
841	ctaaagaaaa	catagagcgc	atgaatgcta	agtataagtt	tgtggttgac	aaaggtagaa
901	gggaattgaa	ttttgaacct	ggagatttgg	tttggttgca	tttgcgaaaa	gaacgatttc
961	ctgatttgag	aaagtctaaa	tagatgccta	gagctgatgg	accatttaaa	gtgttagcaa
1021	agattaatga	gaatgcata	aagattgatt	tgcttgacga	tttcgggggtt	agtcaccacat
1081	ttaacattgc	agatttgaag	ccgtatatgt	ggagaaaaaa	atcgagcttga	gtcgaggatg
1141	actcaaatgc	aagatgggga	ggatgatgag	gacatcaaca	ccatcgatac	atccacgtcc
1201	ccccatatac	agcatgatgg	tcctattarc	cgcgcttgctg	cacgtcaact	aaattatcag
1261	gtgattcttt	cttgagttca	aalttctctg	tctttatacc	tcggagacgc	gtgcaactcgt
1321	gttttactcc	aggaacgtat	ggagaggatc	aaagggaaga	ggattcgcgc	gggttggatt
1381	cggactgcag	ggcagcgcca	acttctgacg	gccgccacga	cttcattgcaa	actccgattt
1441	gggcgtgcaa	gtacttcatg	gaaagcttat	caagtcta		

8/13

FIG. 8

1	gac	ttttt	ct	tgga	tca	aatgt	tcaa	aaatg	ccaa	aacat	gattt	ttggacata
61	tgg	agtgt	at	tgggt	gcatt	catgg	caaaa	actcact	ccg	tgatt	tcg	gcgc
121	tgt	caatt	taa	tgcca	atatt	gggac	acgag	ggtgc	gatgt	tttct	act	gg
181	ttc	aaaaa	aga	gccaaa	acat	gattt	ttt	gga	catatt	ggg	tgatt	tcgt
241	gca	aaact	ca	cttc	gtgatt	cgcg	gcg	ca	acttt	gtca	ttaat	tgca
301	acac	gagaga	gtgc	gatgtt	attct	act	tg	aatca	aaaa	ag	gccaaa	acat
361	gatt	ttt	gga	catatt	ggag	tgatt	tggt	gcgt	tcgt	gg	caaa	ctc
421	gcgc	ggcg	aa	ctttt	gtcaa	ttaat	ttcaa	tatt	gtcata	ttt	tgcccc	aagt
481	atgt	tatt	aa	tcgga	atgaa	aaagt	ttcaa	aggc	accaaa	acat	gattt	ttgacata
541	ggag	tgatt	tt	gggt	gcgttc	gtgg	gaaaaa	ctcact	ttcgt	gatt	tcgcgc	gcgaa
601	gtcatt	ta	at	tg	ccaat	at	tg	gcac	agcgac	gggt	gcgat	

9/13

FIG. 9

1	aagcttcata	ctttgtccgt	ttgaaataag	atttttctgtg	gttgggttcta	gtgtgctcct
61	ggttgtgaaa	aaaaatatca	aaaaataaat	aaaaaaataa	aaaaaaagtt	gtgcctctct
121	cctcagcact	agagagagca	agcaaaaaaa	aaaaaaagagg	atccagcgag	cacaccacac
181	cattccccct	ttgtcaccac	cttccaccac	cggttcttgt	tttcgttgc	gtttgggttt
241	gtgtttttgtg	gcgtcctttg	tgtttagact	cgtcgtctct	acttcgacta	gcctagggacc
301	agctactgta	cactataccc	tctgagcyat	tattcaattt	gctttggcta	acgtgggtttc
361	taattcttct	cttgagcaat	gcttttcagcc	caccaacagc	tccaccacaa	accgactac
421	agcttgacag	gtcttgttgc	tacagcaacg	atacacctcg	tccacggtta	gctaactagt
481	ggtgtgtgtg	tcccctacct	gtgggcaagg	taagaactgg	taagagcttg	tgagacaagc
541	tgcgagtga	gtgagagtga	gcgtctttgca	aaagctacat	cccaatagtt	gtgtagggtt
601	tacattacat	cctccattgt	tttgtttgtcc	ttcatttcta	aacctggca	ggtccagagg
661	ataaagatga	gtagggtgca	tctcattccc	cacgcaccaa	aggcatcatc	caatacttta
721	caaggcaagt	gaagcagcac	acagaaggac	ttgatacaga	tttgcagggtg	acaaatgaga
781	agattggaca	attggagtcc	acgcagatct	ccaccaacac	taagctt	

10/13

FIG. 10

```

1      aagcttctcg ccacaaaaac agaatttttg aagaggaatt tggcgatttt aaatttgccg
61     aaccagcaa agctggggag aaacggatgt aacttttttt gtagatgtcc gtggatatat
121    tataagccta atttcgagtc cgtatgaaaa agatatgact gttttaagga aggtcactcg
181    aatcagggtt ctatttcctt tctagtggcg gcaggataca aggcaactca caaaagctat
241    cggcaacaga tcaaacacaa atagaaagtg atttgggtgg tgcactgcgac ggtggtgacg
301    atggtgatga cggcaaaaatt gatgaagtgg cgatgtgcgg tgggtggcaag aactcgaac
361    tctaattcgc taacactaag accagcaaat taatnaaac gatgcaatcg aaaactcaac
421    aagccttaac taagcagtac taatgaagag gcacaatggt atatatgac accgaaaaac
481    taatctacta tttttttgtt gttgtcttt ctggactata ggaiaaaaaa ataatgacga
541    agggaagggt aaattctctc accgatgaac cagctcttgc taccacctga tgcgaacccg
601    ctagggtttg tggcccgatc ttctgatgag aggtaaagga taactcgatt agatgaagtc
661    gacgttcacg gcccgactac aactgtccaa agacgctgtg ccttagcaac cgatacaccg
721    tctccaatgg tcgccgcgaa ctbtgtgtgc gcgtcaaccc ggcacgaggg gctactcgtcc
781    tgcaagcaat cgaagaacaa gcaagaacaa gtaggacaag cactgaaatt gctagataaa
841    gatgaaagtt tcactatcaa acacaatatg gtgggtttcc gattacaggc agacgcggcg
901    gtttagccaa cacgcgcgct gcgagccggt agcaagaagc tatcttctat catcaaaacc
961    cgcctgtttt tggcggcgac tagaggtata acacaagagg ggaagaacga ccatagggtc
1021   gtgttccaac cctaggacgc gccctaatg gcccacacat ggatacacag cccattgggc
1081   caaaagaggt gacgcagcac catggacaga aaatagtcgg gagtataatg acaattgcgg
1141   ccgctccaga acagatatgg acatgtggct ggatccactt gaaagtagac ttgataagct
1201   t

```

11/13

FIG. 11

```

1  aagcttggtcg cccggggcca ctttagncctt gctgctcttg tgacgcattt tgatgctctc
61  aatgctggag gcaatggtgg gggtaatgat gatgacattg acggagagta cgtcgaggat
121  aatttggagg atgaatatat tgctgacact gaacaagatg atcgagatgc tcgagatcgt
181  cgacgggtac acaacaatcg acgaggtatg ggtggtcgcc gccgacgga ggtacgcaac
241  aatgatgatg ctttttctaa aattaaattt aagattcctc cttttgatgg aaaatatgat
301  cctgatgcgt acctctcttg ggagattgct gttgaccaa aattttgcatt ccatgagttt
361  cctgagagca ctgaggttag agctgctacg agtgagtcca cggattttgc ttgggttttg
421  tggatagaac atggcaagaa aaatcctaata acatgccac aaacttggga tgctttgaaa
481  cgggtcatgc gggctagatt tgttccctct tactatgcac gcgacttgct gaataggttg
541  caacaattga gacaggggtcg gaaactgta gaggaatatt atcaggagtt acaaatgggc
601  ttgcttcgtt gtaattttag gaaactgag gacgctgcca tggctagatt ttgggtggg
661  ttaaaccgag agattttatga catcgtagac tataaagatt acgctaatat gaccgattg
721  ttcatatttg cttgtaaggc tgaaggga catcgtagac tataaagatt acgctaatat gaccgattg
781  aatttttttg caggtaaaac ttcatcatgg cagacacgca ccactcctcc atctagtac
841  actgcttctc catcttccac accacaacc agtcgagcag cacctcctcc atctagtac
901  agtcagcga caaaggctgc tcagccagca ccgagtgttt ctccaatggc atccacaggc
961  cgaatgagag atgttcagtg ccaccgttgc aagggttttg ggcatgtgca gcgtgactgc
1021  cctagcaagc gagtttttgg agtcaaaaaa gatgggtgag actcctctgc tagtgatttc
1081  gatgatgata cacttgcttt gcttgcggct gaccatgcag ataatgagcc accggaagag
1141  cacatttggg ctgcatttgc ggtactat gagagcctca ttgtgcagc ttgctcttagc
1201  gcacaaatgg agaaggcga gcaaatcag cgacacacgt tgttccaaac aaagtgtgc
1261  gtcaaaagac gttgttgcg catgatcatt gatggaggta gctgcaacaa ctgggctagc
1321  agcgagatgg tggagaagct t

```

12/13

## FIG. 12

1	gaatttccttt	gtcacaaagt	gatttactcg	ctgttccttg	tgataaagaa	gagttgtgtg
61	ataatgcttc	acttaaatcc	atgccacaac	tagtgaatga	acatgctatt	cctagtgtat
121	ctttctgtgc	tgattttaag	catgttgttc	acattgctaa	tagagtagag	gaacgtgaat
181	tgacatcttc	tttaataact	ttgggctatg	ttcagtttgc	tgattttcat	gagctcgata
241	atttgaagg	gaaatttatt	gctaagtcctg	atttgccatg	tccaagtaac	gctatttttc
301	atctctttgg	tgaatataat	gatagaggaa	tatatgttgt	gcatagagtt	tacatctgtt
361	cagattttag	acctcctgta	catgtggata	aaacatgcaa	gctagagaga	aatgttattg
421	ctaacaaaat	tgtctcgagt	ttgtcttgtt	ttgattggac	gaaacagggt	gttgtagaat
481	ggtactacgc	aatagagcac	cactatggag	ataataccga	ggacgggtttt	ccattgaaga
541	tacggggagg	tatgatgtga	ccatggctac	acacggatgc	aaactcatgtg	actcacatgc
601	ataggatgaa	cagagaagat	atcgagatca	catcgtccaa	gtgctggaat	ccgattcggc
661	cacactacta	ctactgctga	cttcaaacgg	ccgccgaatt	c	



13/13

## FIG. 13

GAATTCCATCTCGTGTGGCCGGAATCACAGGCACAAGGTGCTGCGTCACCTATA  
ATGGGCGCTATGGGCTTGTAACCTTCATTTGGGACCCCGGCCAGGTGGGGCCCA  
TGTGGGGTGCGCCCAACCAGGTGGAGCACGACCCTAGGCACCCCTAGGTCGT  
CCTCCCACCCCTATATATAGCTAGGTACCCCTTCAGGGTTTCTTGGGTTTGTGATA  
GATTAAAGTTTAGCCATTGCTACTTGCTTGCAGCGCGCGTGTGCGCTAGACCGT  
CCGCTCTGCTTGTCTTCGGAACCCCAACTTATCATTGTATTAAATTCCTATTTG  
CAATATCAGATTGCTTTTATCTTGTCTTGTCTTGTCTTCTTCGATTTGCTTGCAGG  
AATAGGGTTGATCTGCACCGGTAAGATCAACAACCCACGGAGAGGTGTATCGAT  
CGCTAAGGCGCAACACAACGTCTCGTACGGTTGTAGTCGGATCGTCAACGTTTC  
TCCCAAATCGTAGTTATCACAACCTCACCGAAAGATCGGGCCAAACAACCTGCCTT  
GAGTGTGAGAGGAACCTCAGGGTTCATCAGGTGGTATCAGAGCTTTCGTTGCTC  
GGTGAGTTTTTATCTTCTTATAACCAGAAAAATAGCCATACAAAAAAATTCGTAT  
CATTTACCTAGCCATATTGTGCCATTGCATTGTTCTTTTCAGTTTTTGCTTTGTT  
GAATTTTGTGTTGCATCGTCACGTGCTGTTGCTGGTCTTAGCGTCTAGTTTCGTTT  
AGAGTTTCGAGTTCTGGTCACGTTTTGTACCACAGACGTCGCTGCCACCGTGTCT  
TCTGTTGTTGTCGGGACCTACGAAAACGGAATCAGCTCGCTGCCACGGTTCTAT  
TTTTGTAGTTTTTCAGAAGTTTCTGTGCGTTAGTTTTGTAGTTCTTGAGTTGCATC  
TAGAGTTTGCCGATCCTTGTGTGGTTTTGTTTTGGCCATGCCTCTGTCTGTCAC  
GAGAGGAGGAAGGTGAATTACCATATCTGATTTTGGAAAGTAGTCAAGTTTGT  
TGGAAAGATATAGTAGATTGGATTGGTTAAAAATCAGTTTTCCTTTTTATCCCAT  
CACCAAATTCGGCTGCCATCCACTCCACCTCCTGGCCGAGTCCCCTCGCCCCCT  
TTGGCCGAGTCCGACTCCCTCCTCTCTTCCACGATTCCGAGTTGTGTCAAACA  
CCTTGGCGAATTTTTGTTTGGTGTGCGTTTTTCGAGATCTGTTTGGAAAAACGAA  
CCGGCATAAATTCAGCATTCCATTTTTTGTATAATTTTAATTTTGGGTTTAGACT  
TTTACATTTGAGTCCCTGTAAATTTTTATATTTATGTTTGAAGTCCCTGTAACTTA  
CATTAGGTCCCTTGAAGTCAGTTTTTGTAAAAAAATCAAGAAAAAAAGTGAG  
AAAAAAAAGGCAGAAAGGTGCAAAAAAAAGAGAAAAAAAGCCGCACAAA  
AAAAACAGAAAAGAAAAGGAAAAGAAAAAAAGAAAAGAAAAGAAAAGAAA  
GAAAGAAAAGAGAAAATACTGTTATGTTTGAAGCTGAACCTTCATATATCAGACTT  
GTGCTAGAGTTGTTCTTAGTGCTATCTTGTGGTATCGTTTGTGTCTAGGCTCGC  
GTCTCTAGTACGTTTCTAGCCTAGGACCAGCACGGTACTTGACTTTGAACAATTA  
TTCAACTTTGCATTATCTGATTTGAGCATTGCTATTCTTTGCTACATAATTTAA  
GCCTACCCAGAGCTCCACATATTTGATTACAGCCGTACCGCAAGTGTTTGCCAA  
GGCATCGATACATCCAACCTTCATTGGGGTGCTTGGTTGAGTCGGTGTATGTCA  
CCATTCCACTTGCATTGGTAAGATCTTGTAAAGAGCTTGGTTAAAGCTTGAGTG  
TGTGCGATTTTTTGACCTGCCACTACCTAGTAGTTAATAGGAACGCGCATATTTT  
TGTGTATGTTTCTGTTTTCTACTAACAATGGCAGGGATACGCAAGATAATTGG  
GGATAGCTGTGCTCAACATCGACATCTTCGTGCGAGACATGAGGAGGGATCAACA  
TGACCATTATGAGGTAAGTGATGATGTTCTAGGTAAGATCAAATCTGCACTGCC  
TTATTTTCGAGGAACTATGATCCTCGTGCTTACATTAATTGGGAGCTAGCGGT  
TGATAGTGAATTTCAAAGCATGTCTTGTGCGGAGAAACAAAAGGTTATGTGTGC  
CTCTAGTGTTTTAATTAACATGCTTCTAATGATTGGAAACATCTTTGTAGGCAT  
AACAAAATACCACAATCTTGGAAAGACCTGAAACGATATTTTCAGAGATGTTTAT  
GTTCCCATGTATTATGCTGATATCTGCTCAACAACTGCAATGTTTAAACAAG  
ATACCAAAGTGTTACTTCATACTATCATGATATGCATGCTTGTATTACGTTG  
TGGCTTAGATGAATGTGAAGAAGCTACAGAATTGAGGTTTTTACGTGGACTTAA  
CAAAGAAATTCAGGACATGCTTGCTTGTGAAAAGTATAGATCTCTTCTCATTTG  
TTACAACCTTGCTTGAATGCTGAAAGTAAAAATAGAGGAGGATATGAAAAAGAAA  
CACCGTATGCTTTTGCCCTCAATTACTAATCTTTCAGGAAAGTGCATATCAT  
GAAAAGGAGGAGAGAGACATGAAAGAGGCCACCAATTCCATTGTTTACACTCAA  
GTTTCGAGACACCTCCATCATCTAAAGAGGACATCAAAGGTAAAGTAAATGGTAC  
TGAAATTAATCAAGGTGAGTGCAATTGTTAACGAACTAAATTTGTTCACTTTTCAT  
GCAAAAGTAGAGCAACCATTAGTGAACCAAATGCTGGAATTC

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- (i) APPLICANT: Jiang, Jiming  
Dong, Fenggao
- (ii) TITLE OF INVENTION: DNA Sequences Specific to Rice  
Centromeres
- (iii) NUMBER OF SEQUENCES: 8
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: Rockey, Milnamow & Katz, Ltd.
  - (B) STREET: 180 N. Stetson Avenue, 2 Prudential Plaza,  
Suite 4700
  - (C) CITY: Chicago
  - (D) STATE: Illinois
  - (E) COUNTRY: U.S.A.
  - (F) ZIP: 60601
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Floppy disk
  - (B) COMPUTER: IBM PC compatible
  - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER: US
  - (B) FILING DATE:
  - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: Mueller, Lisa V.
  - (B) REGISTRATION NUMBER: 38,978
  - (C) REFERENCE/DOCKET NUMBER: WARF
- (ix) TELECOMMUNICATION INFORMATION:
  - (A) TELEPHONE: 312-616-5400
  - (B) TELEFAX: 312-616-5460

## (2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1478 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (vi) ORIGINAL SOURCE:

(A) ORGANISM: *Oryza sativa* subsp. *indica*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

TCATCAATGA TGGGTTTGTT TTCAGAGCTA ACAAGCTATG CATTCCAGCT AGCTCCGTTT	60
GCTTGTTGTT GTTGACAGAA GCGCATGGAG GCGATTGAT GGGACATTTT GGTGCCAAGA	120
AGACACATGA CATCCTTGCT AGTCATTCTT TTTGGCCACA GATGCGAAGA GATGTTGGCA	180
GGTTCGTTGC TCGCTACGCT ACATGTCAAA AGGCTAAGTC ACGCTTACAT CCACATGGTT	240
TGTATATGCC TCTTCCTGTT CCTACTGTPC CTTGGGAAGA TATTCAATG GATTTTGTGT	300
TAGGATTGCC TAGGACCAAG AGGGGGCGTG ATAGCATTTT TGTGGTTGTG GATCGATTCT	360
CTAAATGGC ACATTTTCATA CCATGTCATA AAAGTATGA TGCTTCTCAT ATCGCTGATT	420
TGTTCTTTTC AGAAATGTT CGCTTGTCAT GTGTGCCAAA CACAATTGTT TCTGATCGTG	480
ACACAAAATT TCTTAGCCAT TTTTGGAGAA CTTTGTGGGC TAAATTGGGG ACTAAACTTT	540
TGTTTTCTAC TACTTGTCAT CCCCAAACTG ATGGACAAAC TGAAGTGGTG AATAGAACCT	600
TGTCTACTAT GCTTAGGGCT GTTTGAAGAA AAATATCAAG ATGTGGGAAG AATGCTTGCC	660
TCATATTGAA TTTGCTTATA ATCGTTCCTT GCATTCTACT ACAAAAAATG TGCCCATTTT	720
AGATTGTGTA TGTTTTGTTG CCTCGTGCTC CAATTGATTT GATGCCTTTA CCATCTTCTG	780
AGAAACTGAA TTTTGATGCG AAGCAACGTG CTGAGTTGAT GTTAAACTG CATGAGACAA	840
CTAAAGAAAA CATAGAGCGC ATGAATGCTA AGTATAAGTT TGCTGGTGAC AAAGGTAGAA	900
GGGAATTGAA TTTTGAACCT GGAGATTGCG TTTGGTTGCA TTTGCGAAAA GAACGATTTT	960
CTGATTGAG AAAGTCTAAA TAGATGCCTA GAGCTGATGG ACCATTTAAA GTGTTAGCAA	1020
AGATTAATGA GAATGCATAT AAGATTGATT TGCTTGCGA TTTCCGGGTT AGTCCACAT	1080
TTAATATTGC AGATTGAAG CCGTATATTG GGAGAAAAAG ATGAGCTTGA GTCGAGGATG	1140
ACTCAAATGC AAGATGGGGA GGATGATGAG GACATCAACA CCATCGATAC ATCCACGTCC	1200
CCCCATATAC AGCATGATGG TCCTATTACC CGCGCTTGTC CACGTCAACT AAATTATCAG	1260
GTGATTCCTT CTTGAGTICA AATTCCTCG TCTTATACC TCGGAGACGC GTGCACTCGT	1320
GTTTTACTCC AGGAACGTAT GGAGAGGATC AAAGGGAAGA GGATTCGCGC GGGGTGGATT	1380
CGGACTGCAG GGCAGCGCCA ACTTCTGACG GCCGCACGA CTTTCATGCA ACTCCGATTT	1440

GGGCGTGCAA GTACTTCATG GAAAGCTTAT CAAGTCTA

1478

## (2) INFORMATION FOR SEQ ID NO:2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 639 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: *Oryza sativa* subsp. *indica*

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GATCTTTTCT ACTGGAATCA AAATGTTCAA AAAATGCCAA AACATGATTT TTGGACATAT	60
TGGAGTGTAT TGGGTGCATT CATGGCAAAA ACTCACTCCG TGATTGCGCG GCGAACTTT	120
TGTCAATTAA TGCCAATATT GGGACACGAG GGTGCGATGT TTTCTACTGG AATCAAAAAG	180
TTCAAAAAGA GCCAAAACAT GATTTTGGGA CATATTGGAG TGTATTGGGG TCGTTCGTG	240
GCAAAACTCA CTTCTGTGATT CGCGCGGCGA ACTTTTGTCA TTTAATGCCA ATATGTGCAT	300
ACACGAGAGA GTGCGATGTT ATTCTACTGG AATCAAAAAG TTCAAAAAGA GCCAAAACAT	360
GATTTTGGGA CATATTGGAG TGTATTGGGT GCGTTCGTGG CAAAACAC TTCGTGATTC	420
GCGCGGCGAA CTTTGTCAA TTAATTCCAA TATGTGCATA TTTGGCCCA AAGTGTGCG	480
ATGTTATTAA TCGGAATGAA AAAGTTCAAA AGGCACCAAA ACATGATTTT TTGACATATT	540
GGAGTGTATT GGGTGCCTTC GTGGGAAAAA CTCACTTCGT GATTGCGCG GCGAACTTTT	600
GTCATTTAAT GCCAATATTG GCACAGCGAC GGGTGCGAT	639

## (2) INFORMATION FOR SEQ ID NO:3:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 827 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: *Oryza sativa* subsp. *indica*

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

AAGCTTCATA CTTTGTCCGT TTGAAATAAG ATTTTCTGTG GTTGGTTCTA GTGTGCTCCT	60
GGTTGTGAAA AAAAATATCA AAAAAATAAT AAAAAAATAA AAAAAAGTT GTGCCTCTCT	120
CCTCAGCACT AGAGAGAGCA AGCAAAAAA AAAAAAGAGG ATCCAGCGAG CACACCACAC	180
CATTCCCCCT TTGTCACCAC CTTCCACCAC CGGTTCTTGT TTTTCGTTGCT GTTTGGGTTT	240
GTGTTTTGTG GCGTCCTTTG TGGTTAGACT CGTCGTCTCT ACTTCGACTA GCCTAGGACC	300
AGCTACTGTA CACTATACCC TCTGAGCGAT TATTCAATTT GCTTTGGCTA ACGTGGTTTC	360
TAATTCTTCT CTTGAGCAAT GCTTTCAGCC CACCAACAGC TCCACCACAA ACCCGACTAC	420
AGCTTGACAG GTCTTGTGTC TACAGCAACG ATACACCTCG TTCCACGGTA GCTAACTAGT	480
GGTGTGTGCTG TCCCTACCT GTGGGCAAGG TAAGAACTGG TAAGAGCTTG TGAGACAAGC	540
TGCGAGTGAA GTGAGAGTGA GCGTCTTGCA AAAGCTACAT CCCAATAGTT GTGTAGGGCT	600
TACATTACAT CCTCCATTGT TTTGTTGTCC TTCATTTCTA AACCATGGCA GGTCCAGAGG	660
ATAAAGATGA GTAGGGTGCA TCTCATTCCC CACGCACCAA AGGCATCATC CAATACTTTA	720
CAAGGCAAGT GAAGCAGCAC ACAGAAGGAC TTGATACAGA TTTGCAGGTG ACAAATGAGA	780
AGATTGGACA ATTGAGATCC ACGCAGATCT CCACCAACAC TAAGCTT	827

## (2) INFORMATION FOR SEQ ID NO:4:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1201 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: *Oryza sativa* subsp. *indica*

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

AAGCTTCTCG CCACAAAAC AGAATTTTGC AAGAGGAATT TGGCGATTTT AAATTGCGG	60
AACCCAGCAA AGCTGGGGAG AAACGGATGT AACTTTTTTT GTAGATGTCC GTGGATATAT	120
TATAAGCCTA ATTTGAGTTC CGTATGAAAA AGATATGACT GTTTTAAGGA AGGTCACTCG	180

AATCAGGGTT CTATTTCTT TCTAGTGGCG GCAGGATACA AGGCAACTCA CAAAAGCTAT	240
CGGCAACAGA TCAAACACAA ATAGAAAGTG ATTTGGTGGG TGA CTGCGAC GGTGGTGACG	300
ATGGTGATGA CGGCAAAATT GATGAAGTGG CGATGTCCGG TGGTGGCAAG AACTCGAAAC	360
TCTAATCGCC TAACACTAAG ACCAGCAAAT TAATAAAACC GATGCAATCG AAAACTCAAC	420
AAGCCTTAAC TAAGCAGTAC TAATGAAGAG GCACAATGGT ATATATGATC ACCGAAAAAC	480
TAATCTACTA TTTTTTTGTT GTTGTCTTTT CTGGACTATA GGAAAAAAA ATAATGACGA	540
AGGGAAGGGT AAATTCTCTC ACCGATGAAC CACGCTCTGC TACCACCTGA TGCGAACCCG	600
CTAGGGTTTG TGGCCCGATC TTTCGATGAG AGGTAAGGGA TAACTCGATT AGATGAAGTC	660
GACGTTACG GCCCGACTAC AACTGTCCAA AGACGCTGTG CCTTACCAAC CGATACACCG	720
TCTCCAATGG TCGCCGCGAA CTGTGGTGC GCGTCAACCC GGCCACGAGG GCACTCGTCC	780
TGCAAGCAAT CGAAGAACAA GCAAGAACAA GTAGGACAAG CACTGAAATT GCTAGATAAA	840
GATGAAAGTT TCACTATCAA ACACAATATG GTGGGGTTCC GATTACAGGC AGACGCGGCG	900
GTTTAGCCAA CACGCGCGCT GCGAGCCGGT AGCAAGAAGC TATCTTCTAT CATCAAAACC	960
CGCTGTTTT TGGCGGCGAC TAGAGGTATA ACACAAGAGG GGAAGAACGA CCATAGGGTC	1020
GTGCTCCAAC CCTAGGAGGC GCGCTTAATG GGCCCAACAT GGATACACAG CCCATTGGGC	1080
CAAAAGAGGT GACGCAGCAC CATGGACAGA AAATAGTCGG GAGTAAATG ACAATTGCGG	1140
CCGCTCCAGA ACAGATATGG ACATGTGGCT GGATCCACTT GAAAGTAGAC TTGATAAGCT	1200
T	1201

## (2) INFORMATION FOR SEQ ID NO:5:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1341 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: *Oryza sativa* subsp. *indica*

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

AAGCTTGTCG CCCGCGGCCA CTTTAGNCTT GCTGCTCTTG TGACGCATTT TGATGCTCTC	60
---	----

AATGCTGGAG GCAATGGTGG GGGTAATGAT GATGACATTG ACGGAGAGTA CGTCGAGGAT	120
AATTTGGAGG ATGAATATAT TGCTGACACT GAACAAGATG ATCGAGATGC TCGAGATCGT	180
CGACGGCTAC ACAACAATCG ACGAGCTATG GGTGGTCGCC GCCGACCGGA GTACGCAAC	240
AATGATGATG CTTTTTCTAA AATTAAATTT AAGATTCTC CTTTGTATGG AAAATATGAT	300
CCTGATGCGT ACCTCTCTTG GGAGATTGCT GTTGACCAAA AATTTGCATG CCATGAGTTT	360
CCTGAGAGCA CTAGAGTTAG AGCTGCTACG AGTGAGTTCA CCGATTTTGC TTTGGTTTGG	420
TGGATAGAAC ATGGCAAGAA AAATCCTAAT AACATGCCAC AAAC TTGGGA TGCTTTGAAA	480
CGGGTCATGC GGGCTAGATT TGTTCTTCT TACTATGCAC GCGACTTGCT GAATAGGTTG	540
CAACAATTGA GACAGGGTGC GAAAAGTGTA GAGGAATATT ATCAGGAGTT ACAAATGGGC	600
TTGCTTCGTT GTAATTTAGA GGAACTGAG GACGCTGCCA TGGCTAGATT TTTGGGTGGG	660
TTAAACCGCG AGATTTATGA CATCGTAGAC TATAAAGATT ACGCTAATAT GACCCGATTG	720
TTTCATCTTG CTTGTAAGGC TGAAAGGAA GTGCAAGGAC GACGTGCTAG TGCCAAGGCT	780
AATTTTCTG CAGGTAAAC TTCATCATGG CAGACACGCA CCACTCCTCC GGCCGGCCGT	840
ACTGCTTCTC CATCTTCCAC ACCCACAACC AGTCGAGCAG CACCTCCTCC ATCTAGTGAC	900
AAGTCAGCGA CAAAGGCTGC TCAGCCAGCA CCGAGTGCTT CTTCAATGGC ATCCACAGGC	960
CGAATGAGAG ATGTTCAAGT CCACCGTTGC AAGGGCTTTG GGCATGTGCA GCGTGACTGC	1020
CCTAGCAAGC GAGTTTGGT AGTCAAAAAC GATGGTGAGT ACTCCTCTGC TAGTGATTTC	1080
GATGATGATA CACTTGCTTT GCTTGCGGCT GACCATGCAG ATAATGAGCC ACCGGAAGAG	1140
CACATTGGGG CTGCATTTGC GGATCACTAT GAGAGCCTCA TTGTGCAGCG TGTCTTAGC	1200
GCACAAATGG AGAAGGCGGA GCAAAATCAG CGACACACGT TGTTCCAAAC AAAGTGTGTC	1260
GTCAAAGAGC GTTGTGCGG CATGATCATT GATGGAGTA GCTGCAACAA CTTGGCTAGC	1320
AGCGAGATGG TGGAGAAGCT T	1341

## (2) INFORMATION FOR SEQ ID NO:6:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 701 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (vi) ORIGINAL SOURCE:

(A) ORGANISM: *Oryza sativa* subsp. *indica*

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
GAATTCCTTT GTCACAAGTT GATTTACTCG CTGTTCTTG TGATAAAGAA GAGTTGTGTG      60
ATAATGCTTC ACTTAAATCC ATGCCACAAC TAGTGAATGA ACATGCTATT CCTAGTGTAT      120
CTTTCTGTGC TGATTTTAAG CATGTTGTTC ACATTGCTAA TAGAGTAGAG GAACGTGAAT      180
TGACATCTTC TTAAATACT TTGGGCTATG TTCAGTTTGC TGATTTTCAT GAGCTCGATA      240
ATTTGAAGGA GAAATTATTT CCTAACTCTG ATTTGCCATG TCCAAGTAAC GCTATTTTTTC      300
ATCTCTTTGG TGAATATAAT GATAGAGGAA TATATTTGGT GCATAGAGTT TACATCTGTT      360
CAGATTTAGA ACCTCCTGTA CATGTGGATA AAACATGCAA GCTAGAGAGA AATGTTATTG      420
CTAACAAAAT TGTCTCGAGT TTGTCTTGTT TTGATTGGAC GAAACAGGTT GTTGTAGAAT      480
GGTACTACGC AATAGAGCAC CACTATGGAG ATAATACCGA GGACGGTTTT CCATTGAAGA      540
TACGGGGAGG TATGATGTGA CCATGGCTAC ACACGGATGC AACTCATGTG ACTCACATGC      600
ATAGGATGAA CAGAGAAGAT ATCGAGATCA CATCGTCCAA GTGCTGGAAT CCGATTCCGGC      660
CACACTACTA CTACTGCTGA CTTCAAACGG CCGCCGAATT C                          701
```

## (2) INFORMATION FOR SEQ ID NO:7:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2863 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (vi) ORIGINAL SOURCE:

(A) ORGANISM: *Oryza sativa* subsp. *indica*

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
GAATTCATC TCGTGTGGCC GGAATCACAG GCACAAGGTG CTGCGTCACC TATAATGGGC      60
CTATGGGCTT GTAAC TTCAT TTGGGACCCC GGCCAGGTG GGGCCCATGT GGGGTGCGCC      120
CCAACCAGGT GGAGCACGAC CCTAGGCACC CCTAGGTCGT CCTCCACCCC CTATATATAG      180
```



CTAGGTACCC	CTTCAGGGTT	TCTTGGGTTT	TGATAGATTA	AAGTTTAGCC	ATTGCTACTT	240
GCTTGCAGCG	CGCGTGTCGC	CTAGACCGTC	CGTCTGCTTG	TTCTTCGGAA	CCCCAACTTA	300
TCATTTGTAT	TAAATTCCTA	TTTGCAATAT	CAGATTGCTT	TTATCTTGTT	CTTGCTTGTT	360
TCTTCGATTT	GCTTGCAGGA	ATAGGGTTGA	TCTGCACCGG	TAAGATCAAC	AACCCACGGA	420
GAGGTGTATC	GATCGCTAAG	GCGCAACACA	ACGTCTCGTA	CGGTTGTAGT	CGGATCGTCA	480
ACGTTTCTCC	CAAATCGTAG	TTATCACAAC	TCACCGAAAG	ATCGGGCCAA	ACAACTGCCT	540
TGAGTGTCTGA	GAGGAACTCA	GGGTTCATCA	GGTGGTATCA	GAGCTTTCGT	TGCTCGGTGA	600
GTTTTTATCT	TCCTATAACC	AGAAAATAGC	CATACAAAAA	AAATTCGTAT	CATTTACCTA	660
GCCATATTGT	GCCATTGCAT	TGTTCTTTTC	AGTTTTTGCT	TTGTTGAATT	TTGTGTTGCA	720
TCGTCACGTC	GTGTTGCTGG	TCTTAGCGTC	TAGTTCGTTT	AGAGTTTTCGA	GTTCTGGTCA	780
CGTTTTGTAC	CACAGACGTC	GCTGCCACCG	TGTCTCTGTT	GTGTCGGGA	CCTACGAAAA	840
CGGAATCAGC	TCGCTGCCAC	GGTCTATTTT	TTGTAGTTTT	CAGAAGTTTC	TGTCGGTTAG	900
TTTGTAGTT	CTTGAGTTGC	ATCTAGAGTT	TGCCGATCCT	TGTGTGGGTT	TGTTTTGGCC	960
ATGCCTCTGT	CGTGCACGAG	AGGAGGAAGG	TGAATTACCA	TATCTGATTT	TGGAAGTAGT	1020
CAAGTTTGTT	TTGGAAGAT	ATAGTAGATT	GGATTGGTTA	AAAATCAGTT	TCCTTTTTAT	1080
CCCATACACC	AAATTCGGCT	GCCATCCACT	CCACCTCCTG	GCCGAGTCCC	ACTCGCCCCT	1140
TTGGCCGAGT	CCGACTCCCT	CCCTCTCTTC	CACGATTCCG	AGTTGTGTCA	AACACCTTGG	1200
CGAATTTTTG	TTTGGTGTGC	GTTTCGAGAT	CTGTTTGCAA	AAACGGAACC	GGCATAAATT	1260
CAGCATTCCA	TTTTTTGTAT	AATTTTAATT	TTGGGTTTAG	ACTTTTACAT	TTGAGTCCCT	1320
GTAAATTTTA	TATTTATGTT	TGAGTCCCTG	TAATCTTACA	TTAAGGTCCT	TGAGTCAGTT	1380
TTTGTTTAAA	AAAATCAAGA	AAAAAAAGTG	AGAAAAAAA	AGGCAGAAAG	GTGCAAAAAA	1440
AAAGAGAAAA	AAAAAGCCGC	ACAAAAAAA	CAGAAAAGAA	AAGGAAAGAA	AAAAAAAAG	1500
AAAAGAAGAA	AGAAGAAAGA	AAGAAAAGAG	AAAATACTGT	TATGTTTGAG	CTGAACTTCA	1560
TATATCAGAG	TTGTGCATGA	GTTGTTCCCTA	GTGCTATCTT	GTGGTATCGT	TTGTGTCTAG	1620
GCTCGCGTCT	CTAGTACGTT	CTAGCCTAGG	ACCAGCACGG	TACTTGACTT	TGAACAATTA	1680
TTCAACTTTG	CATTATCTGA	TTTGAGCATT	TGCTATTCCT	TTGCTACATA	TTTAAGCCTA	1740
CCCAGAGCTC	CACATATTTG	ATTACAGCCG	TACCGCAAGT	GTTTGCCAAG	GCATCGATAC	1800

ATCCAACCTT CATTGGGGTG CTTGGTTGAG TCGGTGTATG TCACCATTCC ACTTGCATTG	1860
GTAAGATCTT GTAAGAGCTT GGTAAAAGC TTGAGTGTGT GCGATTTTTT GACCTGCCAC	1920
TACCTAGTAG TTAATAGGAA CGCGCATATT TTTGTGTATG TTTCCTGTTT TCTACTAACA	1980
ATGGCAGGGA TACGCAAGAT AATTGGGGAT AGCTGTGCTC AACATCGACA TCTTCGTCGA	2040
GACATGAGGA GGGATCAACA TGACCATTAT GAGGTAAGTG ATGATGTTCT AGGTAAGATC	2100
AAATCTGCAC TGCCTTATTT CGAGGGAAAC TATGATCCTC GTGCTTACAT TAATTGGGAG	2160
CTAGCGTTG ATAGTGAATT TCAAAAGCAT GTCTTGTCGG AGAAACAAAA GGTTATGTGT	2220
GCCTCTAGTG TTTTAATTAA ACATGCTTCT AATGATTGGA AACATCTTTG TAGGCATAAC	2280
AAAATACCAC AATCTTGGA AGACCTGAAA CGATATTTCA GAGATGTTA TGTTCCCATG	2340
TATTATGCTG ATATTCTGCT CAACAACTG CAATGTTTAA AACAAGATAC CAAAAGTGTT	2400
ACTTCATACT ATCATGATAT GCATGCTTGT TTATTACGTT GTGGCTTAGA TGAATGTGAA	2460
GAAGCTACAG AATTGAGGTT TTTACGTGGA CTTAACAAAG AAATTCAGGA CATGCTTGCT	2520
TGTGAAAAGT ATAGATCTCT TTCTCATTTG TTACAACTTG CTTGCAATGC TGAAAGTAAA	2580
ATAGAGGAGG ATATGAAAAA GAAACACGCT ATGTCTTTCG CTCCAATTAC TAACTATTTG	2640
CAGGAAGTGC GTAATCATGA AAAGGAGGAG AGAGACATGA AAGAGCCACC AATTCCATTG	2700
TTCACACTCA AGTTCGAGAC ACCTCCATCA TCTAAAGAGG ACATCAAAGG TAAAGTAAAT	2760
GGTACTGAAA TTAATCAAGG TGAGTGCATT GTTAACGAAG TAAATTTGTT CACTTTTCAT	2820
GCAAAAGTAG AGCAACCATT AGTGAACCA AATGCTGGAA TTC	2863

## (2) INFORMATION FOR SEQ ID NO:8:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 745 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Sorghum bicolor

## (vii) IMMEDIATE SOURCE:

- (B) CLONE: pSau3A9

## (x) PUBLICATION INFORMATION:

- (A) AUTHORS: Jiang, Jiming  
Nasuda, Shuhei  
Dong, Fenggao  
Scherrer, Christopher W.  
Woo, Sung-Sick  
Wing, Rod A.  
Gill, Bikram S.  
Ward, David C.
- (B) TITLE: A Conserved Repetitive DNA Element Located in  
Centromeres of Cereal Chromosomes
- (C) JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
- (D) VOLUME: 93
- (F) PAGES: 14210-14213
- (G) DATE: November-1996

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GATCTTTGGA TTGGAACAG TTAAACGGCA ATATGTGCAT GATGATGATT TTAAAGATGT	60
GTTTTTGCAT TGTAAAGGATG GGAAGGCATG GAATAAATTT GTTGTAATG ATGGTTTTGT	120
GTTTAGAGCT AATAAGCTAT GCATTCCAGC TAGCTCTGTT CGTTTGTTGT TGCTACAGGA	180
AGCACATGGA GGTGTTTTGA TGGGACATTT TGGGGCAAAG AAGACGGAGG ACATACTGGC	240
TGGTCATTTT TTTTGCCAA AGATGAGGAG AGATGTGGAG AGATTTATTG CTCGCTGCAC	300
GACATGTCAA AAGGCCAAGT CACGCTTAAA TCCACACGAT TTGAAGCCAT ATTTGGGTGA	360
GGGAGATGAG CTTGAGTCGG GGACGACTCA AATGCAAGAA GGGGAGGATG ATGAGGACAT	420
CAGCACCATC TATACATCCA CACCTACACC CACACCATCG CCAACACCAC TTGGCCCTCT	480
TACTCGTGCC AGTGCCCGTC AACTGAACCA TCAAGTAAGT TTATTCTTAA ACTCTTGTC	540
ATCATATTTA GACAATGGAG ACACGTGCAC TCTTGTTTTG CTTAGGAATG ATGGAGAGGA	600
CCAGAAGCAT AGGGGATTGG TGTAGGCTGG ATTTGGACAG CAAGACAGCA CCAACTTACA	660
ACAACCGCCA TGACTTCATA CAGAGTCCAT TTAAAGCATG CAAGCACTTG ATGGAAACT	720
CGTCAAGTAT ATTTTAGAT GGATC	745

# INTERNATIONAL SEARCH REPORT

International Application No  
PCT/US 00/17535

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/82 C12N5/10 C12Q1/68 A01H5/00

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N C12Q A01H

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, STRAND, WPI Data, PAJ, BIOSIS

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DONG FENGGAO ET AL: "Rice (Oryza sativa) centromeric regions consist of complex DNA." PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES, vol. 95, no. 14, 7 July 1998 (1998-07-07), pages 8135-8140, XP002153043 July 7, 1998 ISSN: 0027-8424 cited in the application	1,22
Y	the whole document --- -/-	2,8-13, 19,21

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

### \* Special categories of cited documents:

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

\*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

\*X\* document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

\*Y\* document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

\*G\* document member of the same patent family

Date of the actual completion of the international search

28 November 2000

Date of mailing of the international search report

11/12/2000

Name and mailing address of the ISA

European Patent Office, F.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl.  
Fax (+31-70) 340-3016

Authorized officer

Holtorf, S

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 00/17535

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>DATABASE EMBL SEQUENCE DATABASE 'Online! 3 December 1998 (1998-12-03) WING, R.A. AND DEAN, R.A.: "a BAC end sequencing framework to sequence the rice genome" XP002153047 accession no. AQ291790</p>	1
Y	<p>NONOMURA K -I ET AL: "Organization of the 1.9-kb repeat unit REC1 in the centromeric region of rice chromosomes." MOLECULAR AND GENERAL GENETICS, vol. 261, no. 1, February 1999 (1999-02), pages 1-10, XP002153044 ISSN: 0026-8925 page 2, left-hand column</p>	2,8-13, 19,21
Y	<p>JIANG JIMING ET AL: "A conserved repetitive DNA element located in the centromeres of cereal chromosomes." PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES, vol. 93, no. 24, 1996, pages 14210-14213, XP002153045 1996 ISSN: 0027-8424 cited in the application abstract, Fig. 4; page 14213, right column, last paragraph</p>	2,8-13, 19,21
Y	<p>DATABASE EMBL SEQUENCE DATABASE 'Online! 6 January 1999 (1999-01-06) DONG, F., ET AL.: "Rice (Oryza sativa) centromeric regions consist of complex DNA" XP002153048 accession no. AF078903</p>	13,19,21
Y	<p>DATABASE EMBL SEQUENCE DATABASE 'Online! 20 July 1998 (1998-07-20) DONG, F., ET AL.: "Rice (Oryza sativa) centromeric regions consist of complex DNA." XP002153049 accession no. AF058902</p>	13,19,21
Y	<p>DATABASE EMBL SEQUENCE DATABASE 'Online! 20 July 1998 (1998-07-20) DONG, F., ET AL.: "Rice (Oryza sativa) centromeric regions consist of complex DNA." XP002153050 accession no. AF058903</p>	13,19,21

-/--

# INTERNATIONAL SEARCH REPORT

Internal Application No  
PCT/US 00/17535

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	<p>DATABASE EMBL SEQUENCE DATABASE 'Online! 20 July 1998 (1998-07-20) DONG, F., ET AL.: "Rice (Oryza sativa) centromeric regions consist of complex DNA." XP002153051 accession no. AF058904</p>	13,19,21
Y	<p>DATABASE EMBL SEQUENCE DATABASE 'Online! 20 July 1998 (1998-07-20) DONG, F., ET AL.: "Rice (Oryza sativa) centromeric regions consist of complex DNA." XP002153052 accession no. AF058905</p>	13,19,21
Y	<p>DATABASE EMBL SEQUENCE DATABASE 'Online! 20 July 1998 (1998-07-20) DONG, F., ET AL.: "Rice (Oryza sativa) centromeric regions consist of complex DNA." XP002153053 accession no. AF058906</p>	13,19,21
A	<p>WO 98 55637 A (ARCH DEV CORP ;PREUSS DAPHNE (US); COPENHAVER GREGORY (US)) 10 December 1998 (1998-12-10) claim 36</p>	1-22
A	<p>WO 89 09219 A (GEN HOSPITAL CORP) 5 October 1989 (1989-10-05) cited in the application the whole document</p>	6,16,17
A	<p>ARAGON-ALCAIDE LUIS ET AL: "A cereal centromeric sequence." CHROMOSOMA (BERLIN), vol. 105, no. 5, 1996, pages 261-268, XP000965076 ISSN: 0009-5915 the whole document</p>	
P,X	<p>DATABASE EMBL SEQUENCE DATABASE 'Online! 31 January 2000 (2000-01-31) SASAKI, T., ET AL.: "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC clone: P0499C11" XP002153054 accession no AP001080</p>	1

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 00/17535

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9855637 A	10-12-1998	AU 7608698 A EP 0986648 A	21-12-1998 22-03-2000
WO 8909219 A	05-10-1989	AU 622900 B AU 3346489 A EP 0338266 A HU 55052 A IL 89453 A JP 3503599 T US 5270201 A ZA 8901809 A	30-04-1992 16-10-1989 25-10-1989 29-04-1991 29-06-1995 15-08-1991 14-12-1993 29-11-1989